Dan Zhang
Bin Wei   *Editors*

Advanced Mechatronics and MEMS Devices II

Springer

Microsystems and Nanosystems

Series editors

Roger T. Howe, Stanford, CA, USA
Antonio J. Ricco, Moffett Field, CA, USA

More information about this series at http://www.springer.com/series/11483

Dan Zhang • Bin Wei
Editors

Advanced Mechatronics and MEMS Devices II

Springer

Editors
Dan Zhang
York University
Toronto, ON, Canada

Bin Wei
University of Ontario Institute of Technology
Oshawa, ON, Canada

Preface

Mechatronics and robotics have been widely used in many arenas, such as manufacturing, medical, and space. With the rapid development of mechatronics and MEMS technologies, a follow-up edition of the book *Advanced Mechatronics and MEMS Devices* is deemed necessary. The aim of *Advanced Mechatronics and MEMS Devices II* is to introduce the state-of-the-art technologies in the field of mechatronics, robotics, and MEMS devices in order to further summarize and improve the methodologies of mechatronics and MEMS devices. Advances made in the past decades will be well described in this book, including mechatronics, robotics, and MEMS-related issues.

We would like to express our deep appreciation to all the authors for their significant contributions to the book. Their commitment, enthusiasm, and technical expertise are what made this book possible. We are also grateful to the publisher for supporting this project and would especially like to thank Marta Moldvai of Springer Science Business Media (USA) for her constructive assistance and cooperation, both with the publishing venture in general and the editorial details. We hope the readers find this book informative and useful.

This book consists of 30 chapters. Chapter 1 presents the theory and application of actuation of elastomeric micro-devices via capillary force technology. Chapter 2 provides insight into the fundamental design, working principles, and practical guidance of MEMS accelerometers. Details of experimental setups, signal conditioning, and data processing are also provided to construct an integrated performance assessment system. Chapter 3 gives an overview of the impact of the change from a focus on analysis, simulation, and modeling combined with outsourcing hardware design to the use of digital fabrication tools allowing a cyclic design process inside the lab, using many examples from various projects, and shares some insights and lessons learned for facilitating and implementing this process. Chapter 4 presents the design of a family of micro-robots capable of object manipulation in a fluidic environment. Chapter 5 discusses how state-of-the-art mobile technologies may be integrated into human-in-the-loop cyber-physical systems and exploited to provide natural mappings for remote interactions with such systems. A demonstrative example is used to show how an intuitive metaphor is

uncovered for performing a balancing task through the teleoperation of a ball and beam test bed. Chapter 6 provides an overview on force/tactile sensor development. By exploiting optoelectronic technology, two tactile sensors that can be used to execute both fine manipulation of objects and safe interaction tasks with humans are designed and realized. Chapter 7 addresses a brief account of issues related to mechanical properties of MEMS. Micro-testing techniques including micro-tensile and micro-fatigue testing along with the hardware are described with typical sample type, shape, and geometry, depicted with diagrams and images. Chapter 8 studies a type of marmot-like rescue robot for mine safety detection and rescuing. The kinematics, maximum stiffness, minimum stiffness, and global stiffness of the head section of the rescue robot are modeled and analyzed. Chapter 9 presents a systematic review of key control schemes for reconfigurable robotic systems, highlighting their benefits and disadvantages, and also reviews the application of these systems at microscale. Chapter 10 gives a detailed overview of MEMS-based sensors and actuators. Chapter 11 proposes a novel sensing approach to in situ particulate material (soot) load measurement in a diesel particulate filter using electrical capacitance tomography (ECT). Chapter 12 provides an overview of three actuation mechanisms that are relevant for biomedical applications of microfluidics. The topics dealt with include dielectrophoresis, acoustophoresis, and magnetophoresis. Chapter 13 reviews a few mechatronic devices designed and used in ASD screening and discusses a few devices used for therapeutic purposes. Chapter 14 conducts a critical and thorough review on vapor/gas sensing properties of a wide range of electrochemically derived metal oxide nano-forms as the sensing layer employing a different device configuration. Chapter 15 develops a wearable blood pressure monitoring system using ultrasound and a microperfusion system using a metal needle with micro-flow channel for measurement of subepidermal biological substances. Chapter 16 discusses the fabrication strategies and materials for the development of physical, chemical, and biosensors. The emerging applications of flexible electronics in wound healing, wearable electronics, implantable devices, and surgical tools, as well as point-of-care diagnostic devices, are also explored. Chapter 17 presents several MEMS devices where the main application is agriculture. Chapter 18 shows the design, fabrication, and testing of a multifunctional MEMS sensor for use in hydraulic systems. The MEMS device is incorporated into a typical fluid power component. Chapter 19 proposes a piezoelectric-actuated rigid nano-needle for single cell wall (SCW) cutting. A fabricated tungsten (W) nano-needle is assembled with a commercial piezoelectric actuator laterally and perpendicularly. Chapter 20 develops a process planning-driven approach for the development of a robotic percussive riveting system for aircraft assembly automation. Chapter 21 introduces photoinduced fabrication technologies for 3D MEMS devices and examines four technologies and their outcome of applications where fabricated feature sizes decrease and resolution increases. Chapter 22 presents a design principle of the OKES by deriving a mathematical model and characterized the OKES performance in terms of working range, positioning accuracy, resolution, linearity, bandwidth, and control effectiveness with the nano-positioning systems. Chapter 23 presents a lab-on-chip microfluidics system for SCM measurement,

related to the force required to drag a single cell and Newton's law of motion inside microfluidics channel. Chapter 24 focuses on the characteristics of micro-manipulation in terms of the types and principles of gripping forces. Chapter 25 discusses three important aspects of inertial microfluidics: fundamental mechanism, microchannel designs, and applications. Chapter 26 provides a detailed overview of the different types of piezoelectric force sensors and the dynamic calibration techniques that have been used to calibrate these sensors. Chapter 27 introduces a magnetically driven micro-robotics system to explain the procedure of developing a magnetic levitation stage and proposes a sensor switching mechanism that combines magnetic flux measurement-based position determination and optical sensor-based position detection. Chapter 28 applies 3D printing molding methods to fabricate a miniature magnetic actuator for an optical image stabilizer, and the application of robust control techniques to actuate the developed miniature magnetic actuators is discussed. Chapter 29 deals with the concept of biofeedback control systems and its structure, and various applicable control methods which are designed to fulfill different system requirements are provided. Chapter 30 develops an inverse adaptive controller design method for the purpose of mitigating the hysteresis effect in the magnetostrictive-actuated dynamic systems.

Finally, we would like to sincerely acknowledge all the friends and colleagues who have contributed to this book.

Toronto, ON, Canada   Dan Zhang
Oshawa, ON, Canada   Bin Wei
February 2016

Contents

Contributors

Mohd Ridzuan Ahmad Department of Control and Mechatronics Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor, Malaysia

Mohsen Akbari Department of Mechanical Engineering, Laboratory for Innovation in MicroEngineering (LiME), University of Victoria, Victoria, BC, Canada

Center for Biomedical Research, University of Victoria, Victoria, BC, Canada

Anas Alazzam Department of Mechanical Engineering, Khalifa University, Abu Dhabi, UAE

A. Albarbar School of Engineering, Manchester Metropolitan University, Manchester, UK

Gursel Alici School of Mechanical, Materials and Mechatronic Engineering, University of Wollongong, Wollongong, NSW, Australia

ARC Center of Excellence for Electromaterials Science, University of Wollongong, Wollongong, NSW, Australia

Alireza Alizadegan Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada

Seyed M. Allameh Northern Kentucky University, Highland Heights, KY, USA

Sohel Anwar Department of Mechanical Engineering, Purdue School of Engineering and Technology, Indiana Univ Purdue Univ Indianapolis, Indianapolis, IN, USA

Christina Barth Department of Mechanical and Aerospace Engineering, University of Virginia, Charlottesville, VA, USA

Vishwas N. Bedekar Department of Engineering Technology, Middle Tennessee State University, Murfreesboro, TN, USA

P. Bhattacharyya Department of Electronics and Telecommunication Engineering, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, West Bengal, India

Zhuming Bi Department of Civil and Mechanical Engineering, Indiana University Purdue University Fort Wayne, Fort Wayne, IN, USA

P.P. Chattopadhyay Department of Metallurgy and Materials Engineering, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, West Bengal, India

Mu Chiao Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada

Jingsong Chu Micromolding Solutions Inc., Brossard, QC, Canada

Andrea Cirillo Dipartimento di Ingegneria Industriale e dell'Informazione, Seconda Università degli Studi di Napoli, Aversa, Italy

Pasquale Cirillo Dipartimento di Ingegneria Industriale e dell'Informazione, Seconda Università degli Studi di Napoli, Aversa, Italy

Tianhong Cui University of Minnesota, Minneapolis, MN, USA

Edwin Dertien Robotics and Mechatronics Group, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands

K. Dutta Department of Electronics and Telecommunication Engineering, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, West Bengal, India

Giuseppe De Maria Dipartimento di Ingegneria Industriale e dell'Informazione, Seconda Università degli Studi di Napoli, Aversa, Italy

Yu Hui Feng Autonomous Systems and Biomechatronics Laboratory, Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada

Baris Fidan University of Waterloo, Waterloo, ON, Canada

Jared A. Frank NYU Tandon School of Engineering, Brooklyn, NY, USA

Toshio Fukuda Institute for Advanced Research, Nagoya University, Nagoya, Aichi, Japan

Faculty of Science and Engineering, Meijo University, Nagoya, Aichi, Japan

Intelligent Robotics Institute, School of Mechatronic Engineering, Beijing Institute of Technology, Beijing, China

Charles Groepper AT Instruments, Cardiff, UK

University of Minnesota, Minneapolis, MN, USA

Gokhan Gungor University of Waterloo, Waterloo, ON, Canada

Y. Haga Graduate School of Biomedical Engineering, Tohoku University, Sendai, Japan

Ragibul Huq Cummins, Inc., Columbus, IN, USA

Vikram Kapila NYU Tandon School of Engineering, Brooklyn, NY, USA

Mir Behrad Khamesee University of Waterloo, Waterloo, ON, Canada

Saud Khashan Department of Mechanical Engineering, Jordan University of Science and Technology, Irbid, Jordan

Jordan University of Science and Technology, Irbid, Jordan

Carl R. Knospe Department of Mechanical and Aerospace Engineering, University of Virginia, Charlottesville, VA, USA

T. Kobayashi Graduate School of Biomedical Engineering, Tohoku University, Sendai, Japan

Daniel T.H. Lai College of Engineering and Science, Victoria University, Footscray, VIC, Australia

ChaBum Lee Department of Mechanical Engineering, Tennessee Technological University, Cookeville, TN, USA

Wee Sit Lee College of Engineering and Science, Victoria University, Footscray, VIC, Australia

Jin Li Complex and Intelligent System Research Laboratory (CISRL), School of Mechanical and Power Engineering, East China University of Science and Technology, Shanghai, China

Perry Y. Li University of Minnesota, Minneapolis, MN, USA

Rui Li College of Engineering and Science, Victoria University, Footscray, VIC, Australia

Weihua Li School of Mechanical, Materials and Mechatronic Engineering, University of Wollongong, Wollongong, NSW, Australia

Yuwen Li School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China

Zhi Li Eindhoven University of Technology, Eindhoven, MB, The Netherlands

Yu Lin Kirchhoff Van-Rob, Aurora, ON, Canada

Bobby Mathew Department of Mechanical Engineering, Khalifa University, Abu Dhabi, UAE

T. Matsunaga Micro System Integration Center, Tohoku University, Sendai, Japan

William W. Melek University of Waterloo, Waterloo, ON, Canada

Iman Mohammad-Rezazadeh Semel Institute for Neuroscience and Human Behavior, UCLA David Geffen School of Medicine, Los Angeles, CA, USA

Hadi Moradi School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

Intelligent Systems Research Institute, SKKU, Suwon, South Korea

Y. Morimoto Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Takeuchi Biohybrid Innovation Project, Exploratory Research for Advanced Technology, Japan Science and Technology, Tokyo, Japan

Pooria Mostafalu Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA

Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA

Ryozo Nagamune Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada

Ciro Natale Dipartimento di Ingegneria Industriale e dell'Informazione, Seconda Università degli Studi di Napoli, Aversa, Italy

Goldie Nejat Autonomous Systems and Biomechatronics Laboratory, Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada

Bradley J. Nelson Institute of Robotics and Intelligent Systems, ETH Zurich, Zurich, Switzerland

Amir Sanati Nezhad Department of Mechanical and Manufacturing Engineering, BioMEMS and Bioinspired Microfluidic Laboratory, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada

Mehdi Nikkhah School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ, USA

Simon Park Schulich School of Engineering, University of Calgary, Calgary, AB, Canada

Roel S. Pieters Institute of Robotics and Intelligent Systems, ETH Zurich, Zurich, Switzerland

Salvatore Pirozzi Dipartimento di Ingegneria Industriale e dell'Informazione, Seconda Università degli Studi di Napoli, Aversa, Italy

D. Serien Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Takeuchi Biohybrid Innovation Project, Exploratory Research for Advanced Technology, Japan Science and Technology, Tokyo, Japan

Kim A. Stelson University of Minnesota, Minneapolis, MN, USA

Stefano Stramigioli Robotics and Mechatronics Group, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands

Chun-Yi Su Concordia University, Montreal, QC, Canada

S. Takeuchi Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Takeuchi Biohybrid Innovation Project, Exploratory Research for Advanced Technology, Japan Science and Technology, Tokyo, Japan

Khalid Hasan Tantawi Department of Career Readiness-Mechatronics, Motlow State Community College, Smyrna, TN, USA

S.H. Teay School of Engineering, Manchester Metropolitan University, Manchester, UK

N. Tsuruoka Graduate School of Biomedical Engineering, Tohoku University, Sendai, Japan

Antonio Valente INESC TEC (formerly INESC Porto) and School of Science and Technology, UTAD University, Vila Real, Portugal

Bin Wei Faculty of Engineering and Applied Science, University of Ontario Institute of Technology, Oshawa, ON, Canada

Fengfeng (Jeff) Xi Department of Aerospace Engineering, Ryerson University, Toronto, ON, Canada

Kaiwen Yuan Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada

Dan Zhang Department of Mechanical Engineering, York University, Toronto, ON, Canada

Jun Zhang School of Mechanical, Materials and Mechatronic Engineering, University of Wollongong, Wollongong, NSW, Australia

Xiuyu Zhang Northeast Dianli University, Jilin, China

Xiaodong Zhang University of Waterloo, Waterloo, ON, Canada

Pan Zhao Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada

Chapter 1
Actuation of Elastomeric Micro Devices via Capillary Forces

Carl R. Knospe and Christina Barth

Abstract Elastomeric materials offer many benefits to MEMS, including smaller device footprints, greater range of motion, and lower required actuation forces. Capillary forces are a novel and promising mechanism to produce motion in the flexible components of microdevices that contain them. Through the phenomenon of electrowetting these forces can be actively controlled, offering a unique capability to achieve large motions with small voltages. The theory and application of this technology is reviewed herein. Experimental results are also presented illustrating the significant potential of the approach for a wide array of applications.

1 Introduction

1.1 Hybrid Microdevices

The advantages of hybrid microdevices with soft elastomeric components are increasingly being recognized [1–3]. Elastomeric components have a much lower modulus of elasticity than the silicon typically employed in MEMS. As a result, elastomeric flexures do not require the large footprint demanded by micromachined silicon flexures to achieve low compliance. Lower modulus also means that hybrid devices require less actuation force to realize a desired deflection. This reduces both the on-chip space needed for actuators and the actuation power. Elastomeric materials also permit much higher strain (>100 %) than silicon, allowing significantly greater linear and rotational motion of components. Additionally, elastomeric

C.R. Knospe (✉) • C. Barth
Department of Mechanical and Aerospace Engineering, University of Virginia, Charlottesville, VA, USA
e-mail: knospe@virginia.edu
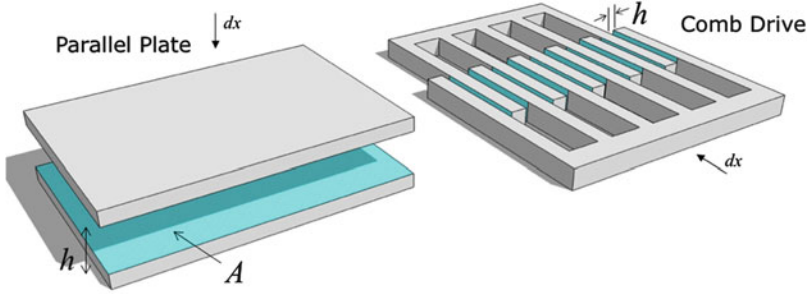
Fig. 1.1 Parallel plate and comb drive electrostatic actuators

materials enable a tight, yet reversible, seal to be formed between surfaces. Thus, on-chip valves and pumps for microfluidic applications can be more easily realized.

A suitable actuation technology is needed, however, to complement the capabilities of elastomeric components, one that full advantage of the high compliance, large deformations, and structural reconfigurability of soft components. While a variety of MEMS actuating technologies are available, including electrostatic, electromagnetic, piezoelectric, and thermal, existing technologies are not up to the task. Each approach has significant disadvantages with respect to one or more of the following performance metrics: force capability, actuator stroke, power required, bandwidth, and ease of integration with elastomeric components [4–7].

To illustrate, let us consider electrostatic actuators, the most commonly used MEMS technology. These can be classified into parallel plate and comb configurations—see Fig. 1.1. In the former, motion is normal to the electrode surfaces and in the later it is parallel with them. As a result, stroke is directly limited by electrode gap in the former. In general, actuators in this family are simple to fabricate using photolithographic micromachining since they do not require materials or elements that are difficult to integrate into microfabrication. Electrostatic actuators typically require voltages much higher than is desirable for MEMS, often greater than 60 V. In many applications, these actuators are very large in comparison to the actuation stroke due to several factors including: (1) limited actuation stroke necessitates significant mechanical amplification to achieve the required range of motion; (2) limitations on device voltage result in large electrode surfaces to achieve the force production required for the application, and for comb drive in particular; (3) the achievable accuracy of lithographically defined features results in relatively large electrostatic gaps, decreasing actuator effectiveness; and (4) "side snap-over" instability limits achievable force and stroke.

In this chapter the electrically controlled capillary forces are investigated for the actuation of elastomeric components in hybrid MEMS. As will be shown, capillary force actuation can achieve significantly greater forces than can be obtained by similarly sized electrostatic actuators. Furthermore, results demonstrate that large actuation strokes can be realized without mechanical amplification.

1.2 Capillary Forces

Consider a liquid droplet in contact with a solid surface and surrounded by a medium, either gas or liquid, referred to as the *ambient*. The contact line surrounding the wetted area of the solid surface has a radius r_c. The interface between the droplet and the ambient has a surface energy, referred to simply as the droplet's surface tension and denoted σ_{al}. The surface energy of the interface between the solid surface and, respectively, the liquid and the ambient will be denoted as σ_{ls} and σ_{as}. The contact angle of the liquid with the solid surface, θ, is determined by these surface energies via $\cos\theta = (\sigma_{as} - \sigma_{ls})/\sigma_{al}$.

Two types of forces are exerted on the solid surface by the droplet: (1) surface tension forces acting at the three phase contact line and tangential to the free interface and (2) Laplace pressure forces acting over the wetted area and normal to the solid's surface [8] (Fig. 1.2). The static pressure difference between the ambient and the liquid is given by Laplace's equation:

$$p = \sigma_{al}\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{1.1}$$

where R_1 and R_2 are the radii of curvature of the liquid's free interface. To illustrate the difference between the surface tension and Laplace forces acting upon a surface consider the case of a short liquid bridge extending between two parallel solid planar surfaces as shown in Fig. 1.3. The radii of curvature at the bridge's waist are $R_1 = r_1$
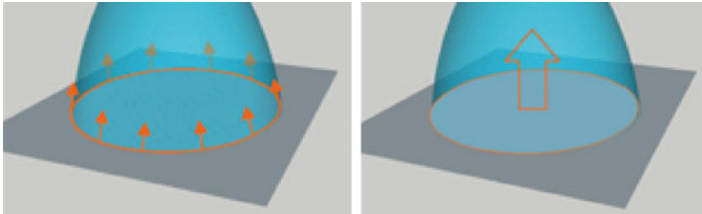
Fig. 1.2 Two types of forces exerted by a liquid droplet upon a solid surface—*Left*: surface tension force acting at the contact line, $F_\sigma = 2\pi r_c\sigma_{al}\sin\theta$; *Right*: Laplace pressure acting over wetted area, $F_p = \pi r_c^2 p$

Fig. 1.3 Geometry of axisymmetric liquid bridge extending between two parallel surfaces of spacing h illustrating radii of curvature of interface
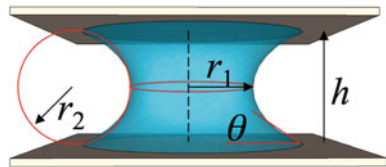
and $R_2 = -r_2 = -h \cos(\theta/2)$. The surface tension force, F_σ, is then found by a line integral of the surface tension over the contact line:

$$F_\sigma = -2\pi \sigma_{\text{al}} \sin \theta \tag{1.2}$$

where the negative sign indicates that this force is acting to bring the solid surfaces closer together. The Laplace pressure is

$$p = \sigma_{\text{al}} \left(\frac{1}{r_1} - \frac{2 \cos \theta}{h} \right) \tag{1.3}$$

The integral of the pressure over the wetted area ($r_1 \sim r_c$) yields the force due to the Laplace pressure:

$$F_p = \pi r_c^2 p = \pi r_c \sigma_{\text{al}} \left\{ 1 - 2 \left(\frac{r_c}{h} \right) \cos \theta \right\} \tag{1.4}$$

The net force exerted upon the solid surface by the drop is then

$$F = \pi r_c \sigma_{\text{al}} \left\{ 1 - 2 \sin \theta - 2 \left(\frac{r_c}{h} \right) \cos \theta \right\} \tag{1.5}$$

The Laplace pressure force will be much larger than the surface tension force if $h \ll r_c$ [8]. Since the Laplace pressure is inversely dependent on the radii of curvature, as the magnitude of the radius R_2 is decreased the magnitude of the Laplace pressure increases rapidly. If properly engineered, capillary forces can provide a highly effective means for actuation in hybrid MEMS.

To illustrate the potential of capillary forces for the deformation of elastomeric microdevices, consider the novel transpiration actuator described by Borno et al. [3] and shown in Fig. 1.4. This passive polymeric microdevice, inspired by the spore dispersal mechanism of ferns, uses the capillary forces of water between ribs to deform the elastic spine connecting the ribs. When not loaded with water, the spine has a circular form of diameter ≈ 1 mm. When water is added, the spine straightens and extends due to the capillary forces acting upon the ribs. As the water evaporates, the spine returns to its original circular form. Between the wetted and dry state the spine's tip moves more than 3.5 mm and rotates more than 330°.

The scaling of capillary forces is particularly well suited for the deformation of such soft structures, inspiring many innovative concepts such as the recently developed capillary origami [9] shown in Fig. 1.5.

Micromechanical systems employing capillary forces do not need to be passive devices like those above. Capillary forces can also be electrically controlled via the phenomenon of electrowetting on dielectric (EWOD). In electrowetting the application of an electrical potential to a conducting liquid placed on a dielectric-covered electrode causes the liquid to spread upon the dielectric surface [10–14]. The apparent contact angle of the liquid on the surface decreases and the surface acts
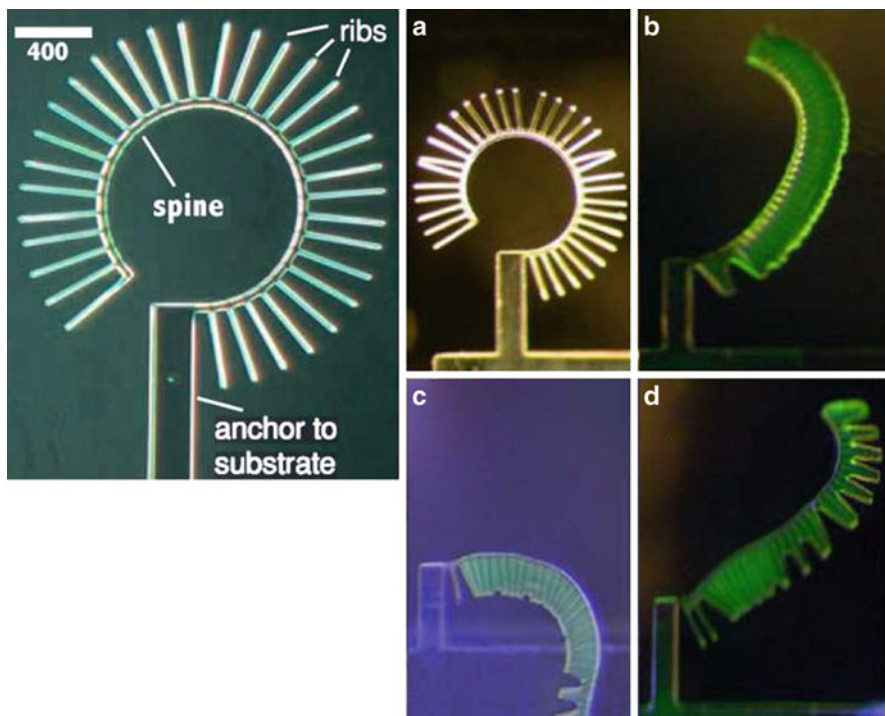
Fig. 1.4 Uncoiling elater—*Left*: microdevice geometry and scale (400 m). *Right*: (**a**) dry device before filling water between fins; (**b**) 0.5 s after filling, device has begun initial deformation due to capillary pressure acting to pull fins together; (**c**) full deflection, 2 s after filling; (**d**) 20 s after filling, evaporation has begun to return the device to its original state. Figure adapted from Borno et al. [3], doi:10.1088/0960-1317/16/11/018 © IOP Publishing. Reproduced with permission. All rights reserved

as increasingly hydrophilic. This change in contact angle results in a change in the curvature of the droplet's free interface and thus induces a change in droplet internal pressure. This change in pressure is the key to capillary force actuation, a novel and promising approach to actuation in MEMS [15–18]. The promise of electrically activated capillary forces for actuation was demonstrated for a conventional MEMS device in [18]. In the experiment a liquid bridge was placed between a fixed plate electrode and a plate electrode supported by micromachined beam flexures. Both electrodes were manufactured from silicon wafers that were covered by a 30 nm oxide film and a 30 nm hydrophobic polymer. The liquid bridge was 100 μm in height and approximately 1 mm in diameter. A maximum deflection of 5 μm was achieved with 40 V applied. This corresponds to an actuation force of over 200 μN. For comparison, a parallel plate electrostatic actuator with the same area and electrode gap would achieve only 3 μN at this voltage. Thus, forces greater than *60 times that of* electrostatic actuation can be achieved via capillary pressure.
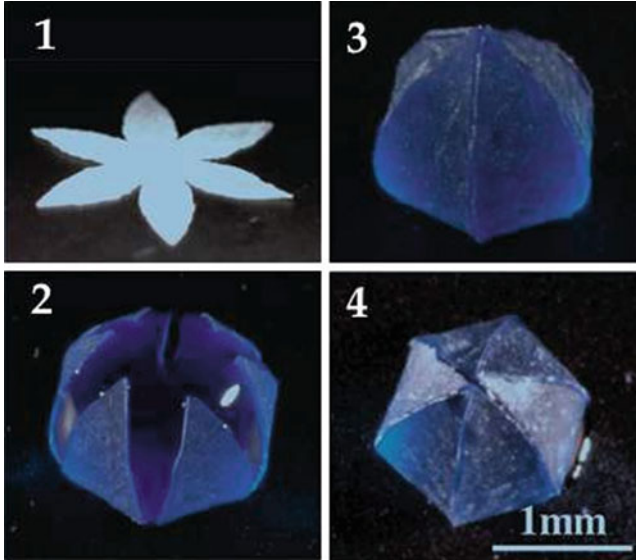
Fig. 1.5 Capillary origami–an elastic planar sheet wraps around a droplet to encapsulate it. (1) sheet before droplet's addition; (2) during the process of encapsulation; (3-4) final configuration. Reprinted figure with permission from [9]. Copyright 2007 by the American Physical Society

In the following section the characteristics of electrowetting that are particularly relevant to the actuation of hybrid microdevices are reviewed.

2 Electrowetting

2.1 Introduction

Consider a typical electrowetting experimental setup (Fig. 1.6): a sessile droplet is surrounded by an ambient gas. The dielectric layer is an oxide that is covered by a thin fluoropolymer topcoat to render the surface hydrophobic and to minimize contact angle hysteresis so as to allow easy movement of the contact line upon the surface. The droplet is an aqueous solution of an inorganic salt. An electric potential is applied between a wire inserted into the drop and the electrode.

Ideally, the change in apparent contact angle in electrowetting is governed by the Young-Lippmann equation:

$$\cos \theta_v = \cos \theta_0 + \frac{\varepsilon_d}{2\sigma_{al}t_d} v_d^2 \qquad (1.6)$$

where θ_v is the apparent contact angle of the liquid with applied field, θ_0 is the native contact angle (without field), ε_d is the permittivity of the dielectric, t_d is the thickness

Fig. 1.6 Typical
electrowetting on dielectric
(EWOD) setup
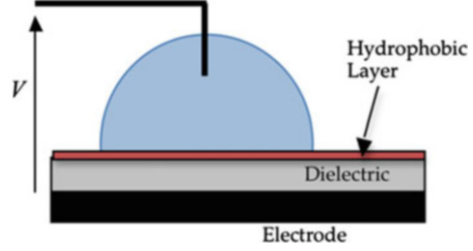
of the dielectric layer, and v_d is the voltage across the layer. Equation (1.6) predicts that perfect wetting (i.e., zero contact angle) will be achieved at a finite voltage. In practice, however, the contact angle will cease to decrease beyond a certain voltage, a phenomenon known as *contact angle saturation*. A variety of mechanisms have been proposed for contact angle saturation, most of which are related to the leakage of charge near the contact line. These include instability of the contact line and expulsion of satellite droplets; charge trapping in the dielectric layer; and electrical breakdown of the dielectric layer [10, 11, 13]. More than one of these mechanisms may contribute to anomalous electrowetting behavior at higher voltages.

2.2 Physics of Electrowetting

Electrowetting is an electromechanical phenomenon in which the applied electric field causes an electrostatic pressure on the liquid surface near the contact line [10]. Thus, it is important to understand the interactions of electric field, ions, and liquid to understand the dynamics of capillary force actuation and how to engineer devices to achieve performance metrics. This discussion will begin with the case of a perfectly conducting liquid. As the liquids typically employed have very high conductivities, this assumption is a reasonable one to begin; its shortcomings will be examined afterwards. When electric potential is applied, it causes an electric field, which results in the migration of ions in the liquid. In the perfectly conducting case, the free charge appears at liquid's interface and screens the electric field from the liquid's interior. An electrostatic pressure is applied to the interface by the interaction of the ions with the electric field induced by the applied voltage [14]. In response to this pressure the conducting liquid is pulled outward causing increased wetting of the substrate. The electrostatic pressure also strongly affects the curvature of the liquid/ambient interface proximate to the contact line. Since the surface energy of the solid/liquid interface is unchanged, the actual contact angle in electrowetting is not altered by the application of voltage—it remains θ_0. The effective contact angle that determines the free interface, however, does change with applied potential and it is typically measured via optical imaging. It is this change in apparent contact angle that is captured by the Young-Lippmann equation, Eq. (1.6). The transition of the liquid profile tangent from native contact angle (θ_0) to apparent contact angle (θ_v) is not easily imaged since the characteristic length scale of the transition is the dielectric thickness. With significant electrowetting the

Fig. 1.7 Helmholtz/diffuse layer model of solvated ions near the contact line during electrowetting (size of ions greatly exaggerated for purposes of illustration)
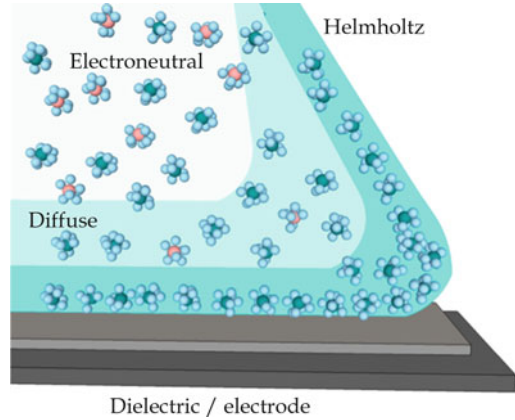
pressure associated with the high concentration of ionic charge and field in this region results in strong local curvature of the gas–liquid interface. The local electric field present in the dielectric layer is dependent on both the electric potential of the liquid and the shape of the interface in this region. In short, application of electric potential results in charge migration, the alteration of the liquid profile, and the redistribution and concentration of the local electric field in the dielectric layer. Thus, the transition region plays a critical role in the mechanisms of charge leakage and electrical breakdown that are thought to result in contact angle saturation and decreased electrowetting performance. In the more realistic case of liquids that do not act as perfect conductors, the transition layer's structure will depend upon the influences of thermal, electrical, and intermolecular forces. Standard models of the liquid/ambient interface consist of a surface layer of oriented water molecules (since dipoles tend to align with field), followed by a Helmholtz lamina of solvated ions, and then a diffuse region of ions in which the potential falls exponentially to that of the bulk (electroneutrality) [19], see Fig. 1.7. We would expect the combined thickness of the Helmholtz and diffuse regions to be in the range of 1–10 nm, depending on ions, solvent, and field. It is clear that these dimensions are not negligible in comparison to the characteristic length scale of the transition region (10–100 nm).

3 Capillary Actuation Dynamics

In examining the actuation dynamics it is useful to consider the time constants of various phenomena in the chain of events that ensue after a change in applied voltage. In a previous paper [16] the author explored this in detail for conventional MEMS devices with capillary actuation via electrowetting. Many of the insights obtained also hold for elastomeric devices and are therefore useful for the present discussion.

When a step in voltage is applied to a hybrid capillary device, there will first be a charging of the dielectric's capacitance. This would occur very quickly with a time constant of about 0.1 μs. Then, the contact angle would change; flow would occur only proximate to the free interface. The time constant for this behavior would be on the order of 10 μs. The local change in interface shape would then set up a pressure differential causing further changes to the free interface, bulk droplet flow, and the beginning of movement/deformation of any participating adjacent elastomeric surfaces. With bulk flow, the liquids free interface would continue to advance and change shape, but at each time instant it would approximately assume an energy-minimizing surface (due to the small time constant associated with capillary waves)—i.e., quasi-static behavior. Typically, the period of bulk flow and solid surface movement would be on the order of 10 ms, slowly ending in the device reaching a new equilibrium and flow ceasing at about 100 ms.

These dynamics are governed by nonlinear differential equations with the nonlinearities arising from four sources: (1) the change in wetted area with the motion of elastomeric surfaces arising from volume conservation; (2) the nonlinear damping introduced by the squeezing of liquid between surfaces; (3) the Lippmann-Young electrowetting equation, and (4) nonlinear elastic behavior associated with large displacements of flexible components. The first source is responsible for any "snap down" (saddle node) bifurcation if it occurs in a hybrid device.

4 Advantages

Capillary force actuation has several advantages for hybrid microdevices.

Large Stroke. Like electrostatic comb drive actuators, capillary force devices produce force through increasing the stored capacitive energy with deflection by increasing the capacitive area. Therefore, the gap between electrodes can remain constant during deflection, unlike the case of electrostatic parallel plate actuators. This feature decouples the thickness of the energy storing capacitance from the actuation stroke. Thus, the thickness can be made as small as practical (so as to improve force capability) without requiring any sacrifice of actuator stroke.

Large Pressure Changes. The pressure change that can be achieved depends on several factors including saturation contact angle, geometry of the free interface, and surface tension of the liquid. The last of these would not be important if contact angle saturation did not occur as it would otherwise cancel out in the governing equations. In all cases, the key factor is the change in capacitance with deformation of the flexible element. The thickness associated with the capacitive energy storage should be as thin as practical so as to maximize this change. With electrowetting actuation this thickness can be much smaller than is achievable with comb drive actuation, 50 nm in comparison to 1000 nm. Furthermore, the permittivity of the capacitive element may be three to ten times greater than the gap in comb drive. Thus, pressures 60 times greater than electrostatic actuation can be achieved, as has been demonstrated in experiments [18].

Application of Pressure to All Surfaces in Contact with Droplet. One key feature of this approach that makes it particularly well suited to the actuation of elastomeric components is that the pressure acts upon all surfaces in contact with the liquid, not just those that contain electrodes, unlike electrostatic actuation. Thus, capillary pressure can be used to deform soft structural elements that do not contain electrodes. This capability can enhance the types of deformations that can be induced and greatly simplify device fabrication. Since the liquid conforms with flexible components as they change shape the pressure will continue to act even after dramatic changes in shape. This capability is made clear by example of capillary origami [9] (Fig. 1.5).

5 Applications

Flexible capillary microdevices have many potential applications ranging from medical care to astronomical optics. Here several potential applications are examined.

Lab-on-a-Chip. Over the past decade rapid progress has been made in the miniaturization of chemical and biological analysis onto microdevices, often referred to as Lab-on-a-Chip [20]. This technology holds many promises including smaller samples, quicker analysis, and parallel processing capability. Critical to this technology is the ability to manage the flow of nanoliter samples in a network of microchannels through flow control devices such as valves and pumps. Active valves technologies today suffer from slow response speed, large power requirements, and bulky off-chip equipment. For example, one common design employs an elastomeric membrane as the valve moving-element, with actuation of this membrane accomplished pneumatically using an external air pump. Similar designs are common for flow generation, including peristaltic and diaphragm pumps [21]. To reduce device size, minimize power consumption, and enhance portability, development of a totally integrated microfluidic device, with flow control integrated on the substrate, is highly desirable. Capillary pressure is an ideal mechanism for the deformation of membranes (or other soft structures) in microfluidic devices, as the electrodes do not need to be integrated into the soft elements.

Adaptive Optics. Adaptive optics systems correct image distortion through the adjustment of the surface shape of a deformable mirror (DM) to alter the light wavefront. AO systems are now common in astronomical imaging and are making inroads into diagnostic retinal imaging [22–26]. Mirror shape change is achieved by electrical signals to arrays of actuators behind the mirror surface. Today, most adaptive optics systems use an array of piezoelectric or electrostatic actuators to deform the DM surface. Actuation requires high voltages and only a limited range of motion (about 1 μm) can be achieved. NSF's Roadmap for the Development of Astronomical Adaptive Optics found that actuation technology for adaptive optics was a "significant technical hurdle" for future telescopes. The report cited large stroke (up to 10 μm) deformable mirrors as a "high priority investment need" [23]. For clinical ophthalmic applications, DMs also need greater stroke for retinal imaging (up to 50 μm) so as to adjust for wavefront errors that occur in the eyes of

the patient population [26]. The required accuracy and stability of motion control for adaptive optics applications is a small fraction of a wavelength and therefore feedback control is typically used in these systems.

Braille Displays. Over 40 million people worldwide are blind and have great difficulty using computers, tablets, e-readers, and cell phones, in spite of the advent of text-to-speech software. Current refreshable Braille displays are both prohibitively expensive and too bulky for these applications [27]. Their complexity also precludes the display of a full page of text, requiring the reader to alter her reading patterns. To achieve a refreshable Braille display with low cost, weight, and power, will require a large stroke actuation technology (300–500 µm) that can be manufactured by high throughput processes such as photolithography and inkjet printing. The required force for holding a raised dot during reading is 0.15 N [27]. However, if actuation is used to release and lower the dot, then it is not necessary that this force level is achieved by the microactuator.

Microendoscopes. Recently miniaturized laser-scanning microendoscopes have been developed for clinical use and biomedical research [28, 29]. These devices require large stroke actuation for x-y scanning and focal depth adjustment. Such devices, if equipped with a significant frame rate, could perform minimally invasive diagnostics in the lungs, urinary tract, and ovaries; and guide therapeutic interventions. This technology could have a profound impact upon health care, eliminating the need for unnecessary, costly, and worrisome biopsies for many patients. Typically, these systems require strokes of 100 µm or more and a scan frequency greater than 10 Hz [29].

6 Example

To illustrate the use of energy approaches for analysis of hybrid devices with capillary actuation an example problem will be examined here. This analysis also illustrates the interplay of surface tension, elastic bending forces, and electrostatic forces in the static behavior of hybrid devices.

Consider the hybrid device illustrated in Fig. 1.8. An elastomeric rectangular membrane of length l and width w is in contact with a droplet that electrowets into a hydrophobic cavity of height h width w. The membrane without any applied forces has curvatures 0 about the x axis and $1/R_u$ about the y axis. The droplet wets the entire underside of the membrane and is pinned along one edge of the membrane and one edge of the cavity structure. The top surface of the cavity is an electrode and the bottom surface is an electrode covered with a dielectric thickness t_d, and permittivity ε_d. The net surface tensions of the top ($i = 1$) and bottom ($i = 2$) surfaces are given by $\sigma_i = \sigma_{ls_i} - \sigma_{as_i}$ where σ_{ls_i} and σ_{as_i} are the surface energies of the liquid/solid and ambient/solid interfaces. The contact angles of the liquid on these two surfaces (without applied voltage) are denoted $\theta_{(1)}$ and $\theta_{(2)}$. The angle of the membrane tangent at its end is denoted ψ. For this analysis the complex liquid interface shape at the device ends (extremes of y) will be approximated by planes as we seek an analytic result; if the width is sufficiently large this approximation will have little impact as the energy associated with end effects will be small. It is assumed that as
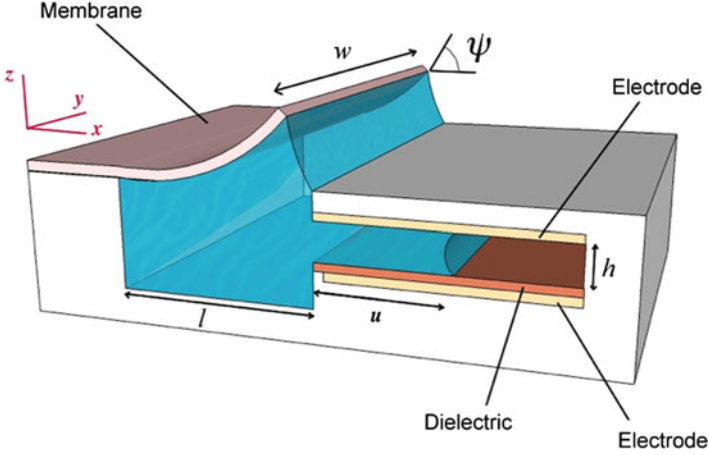
Fig. 1.8 Capillary force actuation of elastomeric membrane in hybrid MEMS device

electrowetting into the cavity occurs the membrane will continue to have uniform curvature with that about the x axis remaining 0. This along with the assumption about end effects yields a single degree of freedom problem. Finally, it is assumed that the variation in area of the liquid/ambient interface within the cavity is small as electrowetting occurs; this is reasonable if the cavity height h is small in comparison to other dimensions. Denote the penetration of the liquid into the cavity by u.

Since the membrane has constant curvature throughout and the length l is constant, the radius of curvature R and the membrane tangent angle ψ are related by

$$R = \frac{l}{\psi} \tag{1.7}$$

From trigonometry, the total volume of liquid is given by

$$V = \frac{w}{2}\left(\frac{l}{\psi}\right)^2 (\sin \psi - \psi \cos \psi) + uwh + V_{\text{dead}} \tag{1.8}$$

where the last term is the dead volume in the device. Taylor series expansion of the first term and linear approximation yields:

$$V = \frac{wl^2}{6}\psi + uwh + V_{\text{dead}} \tag{1.9}$$

This expression is well justified for $0 \leq \psi < 0.9$ rad. Since volume is conserved this expression upon differentiation yields:

$$\frac{du}{d\psi} = -\frac{l^2}{6h} \tag{1.10}$$

The total liquid/ambient surface area can be found via trigonometry and approximated (again, after Talyor series expansion) as:

$$S_{\text{al}} = \left(\frac{wl}{2} + \frac{l^2}{3} \right) \psi + wh \tag{1.11}$$

The total energy of the device is the sum of the capacitive energy, ambient/liquid interface energy, liquid/solid interface energy, and membrane bending energy:

$$E_{\text{total}} = E_{\text{cap}} + E_{\text{al}} + E_{\text{ls}} + E_{\text{bend}} \tag{1.12}$$

The first of these is based on the capacitance of the wetted area of the dielectric:

$$E_{\text{cap}} = -\left(\frac{1}{2} \right) \frac{\varepsilon_d wu}{t_d} v_d^2 \tag{1.13}$$

and the second and third terms are simply the surface tension multiplied by the interface area:

$$E_{\text{al}} = \sigma_{\text{al}} S_{\text{al}} E_{\text{ls}} = (\sigma_1 + \sigma_2)\, wu \tag{1.14}$$

The bending energy can be found in terms of the mean curvature of the membrane with applied capillary pressure (H) and without (H_u)

$$E_{\text{bend}} = \frac{1}{2} K_b (H - H_u)^2 wl \tag{1.15}$$

where K_b is the flexural rigidity, H is the membrane mean curvature, given by

$$H = \frac{1}{2} \left(\kappa_x + \kappa_y \right) \tag{1.16}$$

and the membrane curvatures are $\kappa_x = 0$ and $\kappa_y = 1/R$. For the membrane without applied capillary pressure these quantities are $\kappa_x = 0$ and $\kappa_y = 1/R_u$. Hence, the bending energy is:

$$E_{\text{bend}} = \frac{1}{2} K_b \left(\frac{1}{2R} - \frac{1}{2R_u} \right)^2 wl \tag{1.17}$$

Using Eq. (1.7) this relation can be expressed in terms of the membrane tip tangent angle as:

$$E_{\text{bend}} = \frac{1}{8} K_b \frac{w}{l} (\psi - \psi_u)^2 \tag{1.18}$$

The minimum of the total energy with respect to ψ is the equilibrium for a given voltage. Setting the differential of E_{total} with respect to ψ equal to zero yields the desired expression after some algebra:

$$\psi = \psi_u + 4\frac{l^2}{K_b}\left\{\frac{l}{6h}(\sigma_1 + \sigma_2) - \left(\frac{1}{2} + \frac{l}{3w}\right)\sigma_{\text{al}}\right\} - \frac{1}{3}\left(\frac{l^2}{K_b}\right)\left(\frac{l}{h}\right)\left(\frac{\varepsilon_d}{t_d}\right)v_d^2 \tag{1.19}$$

The membrane tip tangent angle with no voltage applied, $\psi_0\psi_0$, therefore is:

$$\psi_0 = \psi_u + 4\frac{l^2}{K_b}\left\{\frac{l}{6h}(\sigma_1 + \sigma_2) - \left(\frac{1}{2} + \frac{l}{3w}\right)\sigma_{\text{al}}\right\} \tag{1.20}$$

Thus, the no voltage angle ψ_0 is greater than the angle present with no capillary force acting, ψ_0, (i.e., device not loaded with liquid) if

$$\sigma_1 + \sigma_2 > \left(\frac{3h}{l} + \frac{2h}{w}\right)\sigma_{\text{al}} \tag{1.21}$$

Since $\sigma_1/\sigma_{\text{al}} = -\cos\theta_{(i)}$ this condition is:

$$-\cos\theta_{(1)} - \cos\theta_{(2)} > \left(\frac{3h}{l} + \frac{2h}{w}\right) \tag{1.22}$$

which will occur only if the upper and lower surfaces of the cavity are sufficiently hydrophobic. Otherwise, when the droplet is loaded into the structure the capillary pressure will pull the membrane down. Using the contact angle relations Eq. (1.20) can be rewritten as:

$$\psi_0 = \psi_u - \frac{2\sigma_{\text{al}}l^3}{3K_bh}\left\{\cos\theta_{(1)} + \cos\theta_{(2)} + \left(\frac{3h}{l} + \frac{2h}{w}\right)\right\} \tag{1.23}$$

From Eq. (1.19), the membrane tangent angle as a function of voltage is therefore:

$$\psi = \psi_0 - \frac{1}{3}\left(\frac{l^2}{K_b}\right)\left(\frac{l}{h}\right)\left(\frac{\varepsilon_d}{t_d}\right)v_d^2 \tag{1.24}$$

Hence, the degree of membrane actuation at a given voltage is improved by (1) increasing the dielectric permittivity ε_d and decreasing dielectric thickness t_d; (2) increasing the ratio of membrane length l to cavity height h, and (3) decreasing the membrane flexural rigidity K_b.

Consider a device with a PDMS membrane with an 80 μm thickness, 500 μm length, 180 μm width, and membrane tip angle of 40° (without liquid). Suppose that the cavity has height of 40 μm and that the liquid has a contact angle on the upper and lower surfaces of 110°. Furthermore, let the dielectric thickness be 70 nm

and its relative permittivity be 3. After the liquid has been placed in contact with the membrane the membrane tip angle will remain unchanged. As voltage is applied to the device the membrane will uncurl until it achieves a flat configuration ($\psi = 0$) when 10 V is applied.

7 Experimental Results

To illustrate the potential of electrically controlled capillary forces for the actuation of soft structures, consider the actuated membrane prototype developed by the authors (Fig. 1.9). This device has achieved a large actuation stroke while operating with low voltage. The device is composed of a conducting liquid droplet extending between two parallel plates. The upper plate consists of a glass slide in which a 875 μm hole has been laser cut. A 40 μm thick PDMS layer was placed on the top surface of the glass, covering the hole, to form the membrane. The bottom surface of the upper plate was coated with a gold layer to form an electrode. The lower plate
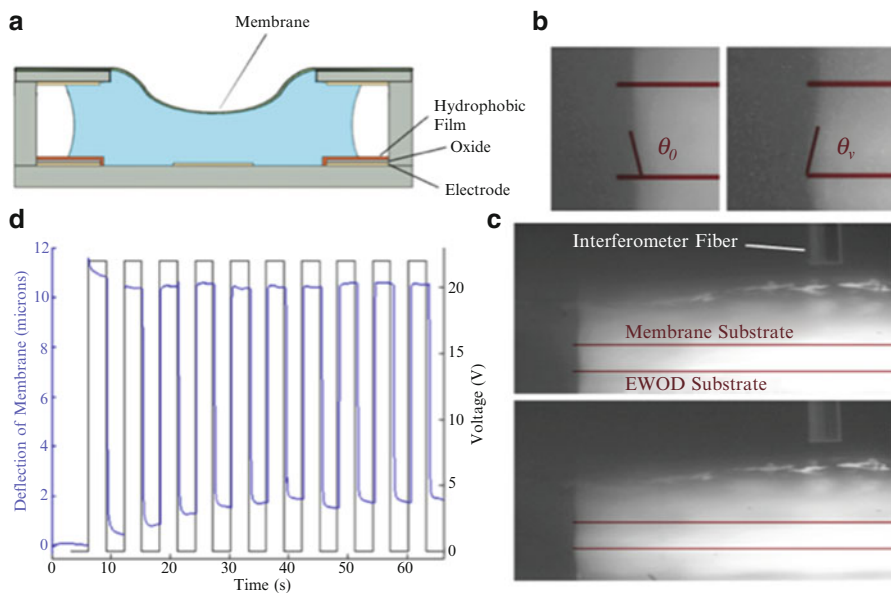
Fig. 1.9 Actuation of an elastomeric membrane via an electrowetting-induced change in capillary pressure. (**a**) Illustration of device configuration; (**b**) close-up images showing change in contact angle on the EWOD substrate, left image is 0 V, right image is 22 V. Droplet is bright due to fluorescence. Location of upper and lower electrodes indicated by *lines*, spacing is 120 mm; (**c**) entire images for 0 and 22 V showing optical fiber used for interferometric measurement of membrane deflection. Multiple droplet images appear due to reflection; and (**d**) membrane deflection and applied voltage during cycling test

consists of a glass slide upon which an aluminum film was deposited. The aluminum was anodized to provide a 44 nm thick dielectric layer of Al_2O_3. This layer was then covered, via spin-coating, with an ultrathin (30 nm) film of hydrophobic fluoropolymer (Cytop). The test device was constructed by depositing a droplet of an aqueous solution (0.1 M Cs_2SO_4 in DI water with 10 μMol/L fluorescein salt) under the membrane and bringing the two surfaces close together, with controlled separation achieved via microbeads spacers. When voltage potential was applied between the electrodes, the droplet electrowetted upon the lower surface, lowering the droplet's capillary pressure, and causing the membrane to deflect downward. The droplet shape was observed via side-view imaging of the droplet's fluorescence under laser excitation. Membrane displacement was measured via a white-light interferometer with light directed by an optical fiber at the center of the membrane. In the test shown, the voltage was cycled between 0 and 22 V. The membrane deflects by over 10 μm. The measured membrane displacement agrees with predicted values to within 10 %, providing an important verification of the actuation principle. Some loss in motion occurs with repeated cycling; this is due to charge injection into the fluoropolymer which then acts as an electret.

8   Conclusion

Capillary pressure offers a unique approach to manipulate microscale structures. It is particularly well suited for the actuation of flexible components since the liquid employed will conform to the elastomeric component during its deformation. Electrowetting provides a means to control the capillary pressure, making this capability available for active devices. A further advantage of this approach is that flexible components need not be electrically conducting. Capillary force actuation can achieve strokes much larger than electrostatic actuation while employing voltages that are significantly lower.

Research is needed on several fronts to further develop this capability. First, the presence of contact angle saturation limits the pressure changes that can be achieved. Even small reductions in the saturation angle would result in substantial improvements in pressure changes that are delivered to flexible surfaces. Second, achieving larger strokes (20–100 μm) is important to the adoption of this technology in a number of applications. This can be accomplished through the optimization of device geometry. Last, better techniques are needed for the fabrication and assembly of these hybrid MEMS. The use of inkjet printing for the loading of droplets into devices is particularly promising.

References

1. Quake S, Scherer A (2000) From micro- to nanofabrication with soft materials. Science 290(5496):1536–1540
2. Roman B, Bico J (2010) Elasto-capillarity: deforming an elastic structure with a liquid droplet. J Phys Condens Matter 22(49):493101
3. Borno B, Steinmeyer R, Maharbiz M (2006) Transpiration actuation: the design, fabrication and characterization of biomimetic microactuators driven by the surface tension of water. J Micromech Microeng 16(11):2375–2383
4. Bell D, Lu T, Fleck N, Spearing S (2005) MEMS actuators and sensors: observations on their performance and selection for purpose. J Micromech Microeng 15:153–164
5. Beyeler F, Neild A, Oberti S, Bell D, Sun Y, Dual B, Nelson B (2007) Monolithically fabricated microgripper with integrated force sensor for manipulating micro objects and biological cells aligned in an ultrasonic field. IEEE J Microelectromech Syst 16(1):7–15
6. Bordatchev E, Zeman M, Knopf G (2005) Electro-thermo-dynamic performance of a micro-gripping system. In: Proc. IEEE International Conference on Mechatronics and Automation, pp 1848–1853
7. Thielicke E, Obermeier E (2000) Microactuators and their technologies. Mechatronics 10:431–455
8. Nezamoddini SA (2008) Capillary force actuators. Dissertation, University of Virginia
9. Py C, Reverdy L, Doppler J, Bico B, Roman, Baroud C (2007) Capillary origami: spontaneous wrapping of a droplet with an elastic sheet. Phys Rev Lett 98:156103. doi:http://dx.doi.org/10.1103/PhysRevLett.98.156103
10. Mugele F, Baret JC (2005) Electrowetting: from basics to applications. J Phys Condens Matter 17:705–774
11. Mugele F (2009) Fundamental challenges in electrowetting: from equilibrium shapes to contact angle saturation and drop dynamics. Soft Matter 5:3377–3384
12. Pollack M, Shenderov A, Fair R (2002) Electrowetting-based actuation of droplets for integrated microfluidics. Lab Chip 2:96–101
13. Vallet M, Vallade M, Berge B (1999) Limiting phenomena for the spreading of water on polymer films by electrowetting. Eur Phys J 11:583
14. Mugele F, Buehrle J (2007) Equilibrium drop surface profiles in electric fields. J Phys Condens Matter 19:375112
15. Knospe C, Nezamoddini SA (2010) Capillary force actuation. J Micro-Nano Mechatronics 5(3):57–68. doi:10.1007/s12213-009-0023-4
16. Knospe C, Haj-Hariri H (2011) Capillary force actuators: modeling, dynamics, and equilibria. Mechatronics 22(3):251–256, Special Issue on Micro-Nano-Mechatronics
17. Knospe C (2011) Capillary force actuation: a mechatronic perspective. In: Eleftheriou E, Moheimani SOR (eds) Control technologies for emerging micro and nanoscale systems, vol 413, Lecture notes in control and information sciences. Springer, Berlin, pp 201–218
18. Wang H, Gaskins J, Knospe C, Reed M (2012) The capillary force actuator: design, fabrication, and characterization. In: 25th IEEE International Conference on Micro Electro Mechanical Systems, Paris, 29 Jan–2 Feb
19. Bockris J, Reddy A (1974) Modern electrochemistry. Plenum Press, New York
20. Oosterbroek R, Berg A (2003) Lab-on-a-chip: miniaturized systems for (bio)chemical analysis and synthesis. Elsevier, Amsterdam
21. Unger MA, Chou HP, Thorsen T, Scherer A, Quake S (2000) Monolithic microfabricated valves and pumps by multilayer soft lithography. Science 288:113–116
22. Fernández B, Fernández B, Bouchti M, Kubby J, Kubby J (2013) High-stroke, high-order MEMS deformable mirrors. J Micro/Nanolithogr MEMS MOEMS 12(3):033012
23. A Roadmap for the Development of Astronomical Adaptive Optics (2008) National Optical Astronomy Observatory. Available via http://www.noao.edu/. Accessed 15 Oct 2014

24. MacIntosh B, Graham J, Oppenheimer B, Poyneer L, Sivaramakrishnan A, Veran J (2006) MEMS-based extreme adaptive optics for planet detection. In: Proceedings SPIE, vol 6113
25. Bifano T, Cornelissen S, Bierden P (2010) MEMS deformable mirrors in astronomical adaptive optics. In: 1st AOELT Conference, EDP Sciences
26. Wu X, Li H, Yao L, Ou H, Pang C (2012) Large stroke actuators and mirror devices for ocular adaptive optics. In: Proc. SPIE 8253, MEMS Adaptive Optics VI. doi:10.1117/12.905697
27. Bar-Cohen Y (2009) Electroactive polymers for refreshable Braille displays. SPIE Newsroom. doi:10.1117/2.1200909.1738
28. Jung W, Tang S, McCormic D et al (2008) Miniaturized probe based on a microelectromechanical system mirror for multiphoton microscopy. Opt Lett 33:1324–1326
29. Domke J, Rhee C, Liu Z, Wang T, Oldham K (2011) Amplifying transmission and compact suspension for a low-profile, large-displacement piezoelectric actuator. J Micromech Microeng 21(6):067004. doi:10.1088/0960-1317/21/6/067004

Chapter 2
MEMS Accelerometers: Testing and Practical Approach for Smart Sensing and Machinery Diagnostics

A. Albarbar and S.H. Teay

Abstract Micro-Electro Mechanical Systems (MEMS)-based sensing elements are gaining wider acceptance and adoption for static and dynamic (mobile) applications. Recent increase in demands for reliable wireless sensing nodes has necessitated seeking alternatives to expensive conventional accelerometers to perform multi-control and monitoring tasks. Owed to their size and cost, MEMS accelerometers is one of the alternative options.

This chapter provides insight into the fundamental design, working principles and practical guidance to MEMS accelerometers. Details of experimental set-ups, signal conditioning and data processing are also provided to construct integrated performance assessment system. Performance assessments are carried out using sinusoidal excitations, impulsive (hummer testing) and random excitations. Subsequently, calculations and comments on frequency response functions, signal-to-noise ratios and phase distortions are outlined. Finally, guidelines to practical adoption of MEMS accelerometers such as packaging, establishing smart vibration sensing nodes and extraction of condition-related information are given.

Keywords MEMS accelerometers design and testing • Performance assessment of MEMS accelerometers • Vibration measurement using MEMS sensors • Smart sensor design and implementation • Machinery-condition monitoring

1   Introduction

Vibration analysis is one of the most usable methods in machinery-condition monitoring. It plays a significant role in the dynamic qualification of newly designed structural components, prediction of faults and structural aging-related problems, and several other structural dynamics studies and diagnosis [1–3]. One reason for

A. Albarbar (✉) • S.H. Teay
School of Engineering, Manchester Metropolitan University, All Saints Building,
All Saints, Manchester M15 6BH, UK
e-mail: a.albarbar@mmu.ac.uk

its wide use is its capacity to monitor vibrating machines without interrupting normal operations. In addition, the vibrating mechanisms of most machineries and structures are fundamentally well known, giving rise to the possibility of detecting many faults in accordance with the characteristics of the vibration responses. Furthermore, the progress of vibration signal processing techniques, computing capabilities, reliable performance of vibration instrumentation such as wide-band transducers and portable analysers has caused this technique to be extensively used around the world.

As an example, in machinery-condition monitoring, accelerometers are often used to measure their vibration (acceleration) signals, which may lead to the detection of any deviation from normal signatures. Vibration waveforms are interpreted and processed in a variety of ways such as peak values and variance of the signal in the time domain, and power spectral analysis in the frequency domain [4].

Vibration has traditionally been sensed using piezoelectric accelerometers. These are accurate and reliable, but have some inherent problems: they are difficult to mass-produce, and they have high source impedance, which means that their signals must be very carefully amplified. Moreover, the use of the traditional piezoelectric accelerometers for simultaneous multiple data collection points was considered to be impractical; this is mainly because of their cost as well as the costs of the associated electronic signal conditioning units. Other types such as piezoresistive accelerometers have limited resolution and can be used only for low and medium frequencies. Low resolution is also a disadvantage of electrodynamic accelerometers. The capacitive type accelerometers have low resolution and are fragile. Commercially, several types of accelerometers manufactured by many well-known manufacturers are available all over the world. In vibration-measurement filed tests, accelerometers of mainly Integrated Circuit Piezoelectric (ICP) type, are found not to be able to measure the non-stationary impulsive responses of structures accurately all the time. However, confidence in measuring impulsive signals using the charge-type piezoelectric accelerometers is very good [4, 5].

Obviously, the quality of the vibration-based diagnosis and/or identification of vibration-related problems mainly depends on the measured responses using accelerometers. Therefore, good performance and high reliability of the commercially available accelerometers is very important. Additionally, the use of conventional accelerometers for multiple data collection points may increase the complexity of monitoring system because of the associated electronic units that are externally connected to the accelerometers. Consequently, the need for cheaper and more reliable devices is well recognised.

With increasing demands for wireless sensing nodes used for assets control and condition monitoring; the need for alternatives to those expensive conventional accelerometers in vibration measurements has risen. Micro-Electro Mechanical Systems (MEMS) accelerometers are one of the available options because of their small size, newer technology and low cost, e.g. MEMS accelerometer may be within 10 % of the cheapest commercially available conventional accelerometers that come with a signal conditioning unit [4, 5]. According to Yole's Development report, the total market for MEMS is expected to exceed $20 Billion in 2020.

The average price of MEMS accelerometers across all applications decreases, from an average of $2.50 in 2004 to less than $1.70 in 2015, with consumer applications driving the price down [6]. MEMS technology is used in some sectors such as automotive industry for measuring pressure, temperature and air bags systems. A few earlier researches compared the performance of MEMS accelerometers with the conventional ones. The main differences are mainly in actuation frequencies, amplitudes and phase shifts. In general, the selection of accelerometers is based on their technical specifications and the measurement requirements. The main technical specifications that decide the use of accelerometers are as the following:

(a) **Sensitivity** is the ratio of the electrical output of the accelerometer to its mechanical input. The output usually is expressed in terms of voltage per unit acceleration. The specification of sensitivity is sufficient for instruments, which generate their own voltage, independent of an external power source. The sensitivity of an instrument requiring an external voltage usually is specified in terms of output voltage per unit of voltage supplied to the instrument per unit displacement, velocity, or acceleration, e.g. milli-Volts per g of acceleration.

(b) **Frequency range** is the operating frequency range over which the sensitivity of the transducer does not vary more than a stated percentage from the rated sensitivity. The range may be limited by the electrical or mechanical characteristics of the transducer or by its associated auxiliary equipment.

(c) **Amplitude limit** specifies the maximum range of acceleration that can be measured by the accelerometer.

(d) **Shock limit** represents the maximum level of acceleration the accelerometer can withstand without any damage to the unit.

(e) **Linearity** indicates the accuracy of the measured acceleration amplitude within the corresponding frequency range.

(f) **Natural frequency** which is indirectly indicative of the measuring frequency range. In general, a higher natural frequency allows a larger measuring frequency range of an accelerometer.

However, the use of vibration analysis in condition-monitoring is based on three key points:

1. A frequency component identifies the basic problem.
2. The amplitudes of this component and its harmonics indicate the severity of the problem.
3. Phase relationships are used to distinguish between looseness and eccentricity (fault diagnosis).

While a vibration spectrum can reveal much about the ranges of motion and flexibility (resonance) of a machine, the frequency domain does not yield all the answers. The time domain is the only place we can identify peak and peak-to-peak amplitudes of each cycle, phase relationship between signals, and the presence of such distinctive characteristics as truncated waveforms, pulses and modulation. The calibration procedure adopted, generally uses the sinusoidal vibration generator (shaker) with varying frequencies and amplitudes to characterise the accelerometer

being calibrated. This is achieved by comparing its measured responses with other well-calibrated accelerometers. Accuracy of sinusoidal response measurement satisfies the calibration procedure requirements. However, accelerometers are in general used for measuring periodic signals (e.g. sinusoidal), as well as impulsive and random signals.

2 Types and Design of MEMS Accelerometer

Generally, MEMS accelerometers are divided into two groups: piezoresistive MEMS accelerometers and capacitive MEMS accelerometers.

2.1 Piezoresistive MEMS Accelerometers

Piezoresistive MEMS accelerometer is also known as strain gauge MEMS accelerometer. Unlike piezoelectric-typed sensor, piezoresistive MEMS accelerometer changes in resistance across the piezo material when stress is applied. The history of piezoresistive MEMS accelerometer could be found in late 1970s. Roylance et al. have successfully developed a first prototype of silicon-based strain gauge accelerometer at Stanford University [7]. The mechanism of piezoresistive MEMS accelerometer is simple. The first design of silicon piezoresistive MEMS accelerometer consists of a frame, a proof mass and a thin beam as shown in Fig. 2.1. The strain gauge is placed on the top layer of the beam, forming a half bridge connection. When acceleration is applied to the sensor, the normal force causes the beam to bend, leading to a change in resistance of the strain gauge. The half bridge configuration gives proportional output to the acceleration. The proposed design has become a fundamental topology of piezoresistive MEMS accelerometer. Multiple beam-mass system has been introduced to improve the accelerometer's merit of figures [8]. Technically, piezoresistive MEMS accelerometer provides good sensitivity. However, their low tolerance on ambient temperature conditions and self-heating characteristics have been the development bottleneck [9]. Benefitting from recent micro-fabrication technology, such difficulties could be overcome and nowadays, there are relatively robust piezoresistive MEMS accelerometers available in the market [7].

Using similar design techniques, higher number of axis, e.g. three-axis piezoresistive accelerometers, could also be manufactured as shown in Fig. 2.2. Four beams with eight piezoresistors configuration are integrated to sense acceleration in three different directions. This configuration could be visualised in an equivalent Wheatstone bridge topology as shown in Fig. 2.2c. The parallel accelerations along x, y and z-axis introduce beam deformation and cause compressive and tensile stress
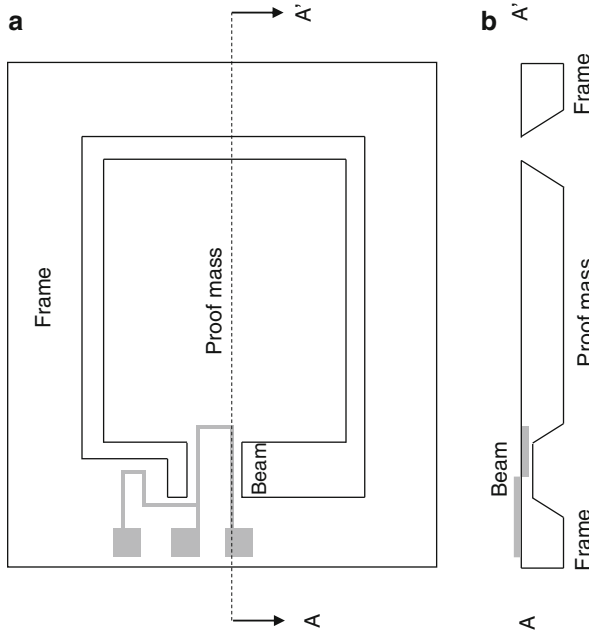
Fig. 2.1 Schematic diagram of piezoresistive MEMS accelerometer (**a**) top view, (**b**) cross-sectional view [7]

in each piezoresistor. The corresponding changes are outlined in Table 2.1. One additional rationale of using Wheatstone bridge is to nullify the defects caused by temperature drifts in piezoresistors [7, 8].

For packaging design, the piezoresistive MEMS accelerometer is wafer encapsulated with etched glass caps, preventing over-range protection and fluidic squeezed-film damping. To ensure better measurement experience, the transfer function between sensor surface and glass gap can be evaluated using their damping coefficients based on mass-spring-damper mechanical system analysis [8].

2.2 Capacitive MEMS Accelerometer

Silicon design of capacitive MEMS accelerometer allows low cost mass production due to the mature technology on surface micromachining. ADXL series from Analog Devices is one of dominant capacitive MEMS accelerometers in the market [4]. Application of capacitive MEMS accelerometer include smartphone devices, smart sensor system and low cost embedded monitoring system. Essentially, the schematic diagram of capacitive MEMS accelerometer is illustrated in Fig. 2.3. The movement of the proof mass due to an acceleration of the device changes the capacitance value between electrodes on stationary fingers. This difference in
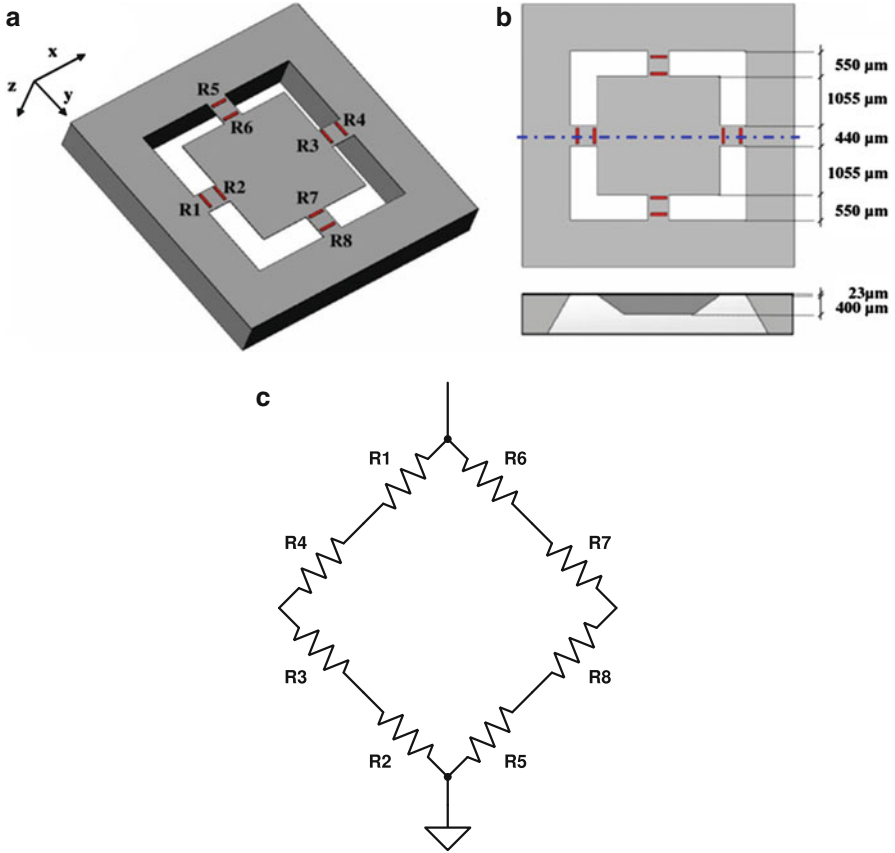
Fig. 2.2 Structural schematic of three axis piezoresistive MEMS accelerometer [9] (**a**) model view, (**b**) top and cross-sectional view, (**c**) equivalent Wheatstone bridge model

capacitance can be used to measure acceleration. An equivalent electronic circuit incorporated with the capacitive model in Fig. 2.3 is shown in Fig. 2.4. The part inside the rectangular box represents which its output is the differential capacitance value from electrodes. The equivalent circuit contains signal source, a buffer amplifier, a filter, a synchronous demodulator and electro-mechanical feedback loop with a feedback amplifier. This model has the sensitivity of 20 mV/g for an operational range of 0–50 g [9, 13]. To improve accuracy, symmetric design and differential sensing are able to reduce the effect of thermal mismatch to a minimum and linearises the capacitance different to acceleration relationship. In modern design, Zhou et al. have proposed an MEMS capacitive accelerometer with fully symmetrical double-sided H-shaped beam, of which the sensitivity of the device is 0.24 V/g with nonlinearity of 0.29 % over the range of 0–1 g [14].

Optimisation problem is always a concern in designing a robust capacitive MEMS accelerometer. General design criteria include manufacturing costs,

Table 2.1 Piezoresistance value corresponding to acceleration direction

Acceleration direction	R1	R2	R3	R4	R5	R6	R7	R8
X axis	Decrease	Increase	Decrease	Increase	Unchanged	Unchanged	Unchanged	Unchanged
Y axis	Unchanged	Unchanged	Unchanged	Unchanged	Decrease	Increase	Decrease	Increase
Z axis	Decrease	Increase	Increase	Decrease	Decrease	Increase	Increase	Decrease
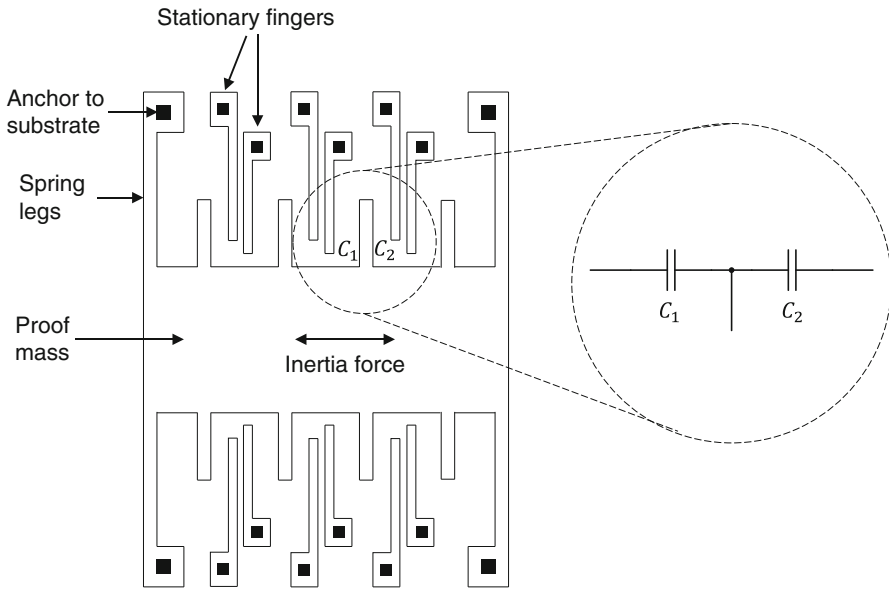
Fig. 2.3 Schematic diagram breakdown respective circuit model of capacitive MEMS accelerometer [10]
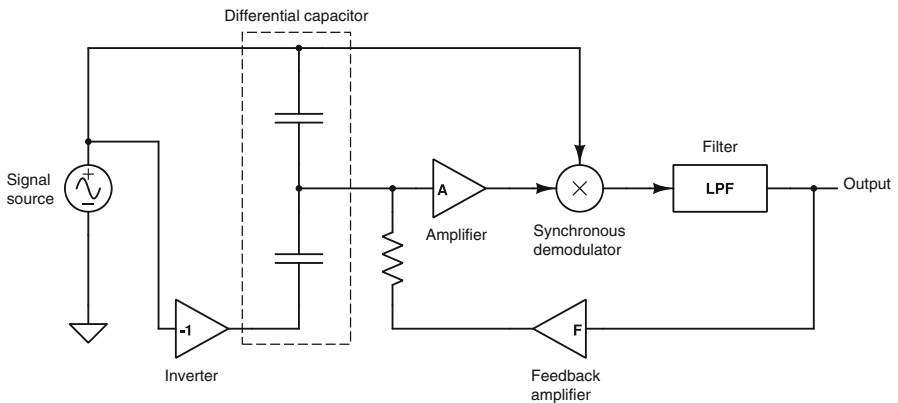
Fig. 2.4 Equivalent circuit of the integrated MEMS accelerometer [11, 12]

full-scale range, structural geometry and threshold acceleration [15]. Due to the complexity of the mechanism, the performance of capacitive MEMS accelerometer is still far from trivial. Various techniques have been carried to study the variables of capacitive MEMS accelerometer to optimise its performance and robustness [16, 17].

3   Testing and Calibration

Calibration is carried out on a variety of equipment in different situations to ensure that the equipment will produce results, which meet pre-defined operating criteria. In the context of calibration of devices (MEMS accelerometers), a calibration involves determining the performance of the sensor in reference to a known standard. Adjustment of the overriding system may be a part of a calibration, but not always.

In practical applications, accelerometers are usually used for measuring the periodic (sinusoidal, sweep-sine, step-sine, multi-sine, etc.), impulsive and random signals. Hence, these tests are usually carried out on a similar test set-up to that shown in Fig. 2.5, and subsequently the outputs of MEMS accelerometers are compared with well-calibrated accelerometers.

3.1   Performance Evaluation

In order to assure meaningful outputs, the performance of MEMS accelerometers have to be critically assessed and calibrated. Albarbar et al. have conducted a comprehensive study on performance of MEMS accelerometer for machinery-condition monitoring [4, 5]. Three MEMS accelerometers (denoted as A, B and C) have been validated by comparing their performance with a well-calibrated, conventional accelerometer and subsequently were utilised to measure the vibration signals of CNC machine in real industrial environment. Their respective technical specification is listed in Table 2.2. The tested devices output frequency response functions (FRS) and phase of the outputs of the tested devices are shown in Fig. 2.6a, b, respectively.
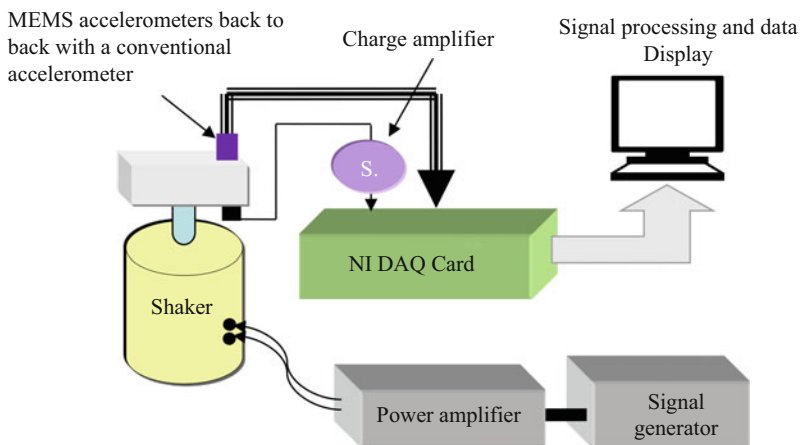
Fig. 2.5 MEMS accelerometer's calibration experimental set-up

Table 2.2 Technical specifications of accelerometers

	Conventional accelerometer	MEMS (A)	MEMS (B)	MEMS (C)
Sensitivity (V/g)	0.1	Supply voltage/7	Supply voltage/5	Supply voltage/5
Frequency range (Hz)	1–2000	1–6000	1–10,000	1500
Amplitude limited (g)	±50	±5	±3	±3
Linearity	Less than ±5 %	±1 %/1 kHz	±1 %/5 kHz	±1 %/1 kHz
Shock limit (g)	5000	250	1000	100

As anticipated, there were clear differences between the tested accelerometers due to their design, sensitivity and noise contents of their outputs. For more details, the reader is advised to see [4, 5].

3.2 Calibration of MEMS Accelerometers

In this section, the calibration is performed using recently developed MEMS accelerometer (ADXL001-70) (http://www.analog.com/media/en/technical-documentation/data-sheets/ADXL001.pdf). An accelerometer calibrator type-4291 by Brüel & Kjær is used to generate reference mechanical vibration at $1\,g$ ($10\,\mathrm{ms}^{-2}$) peak $\pm 2\,\%$ with 79.6 Hz. A conventional Brüel & Kjær piezoelectric accelerometer was used as a reference and a commercial data acquisition system by National Instruments recorded the outputs. Both accelerometers are attached on the mechanical shaker, which excites at different frequency and amplitude. A photo of the calibration set-up is shown in Fig. 2.7. The calibration results are shown in Table 2.3 and in Fig. 2.8.

In the other test, the mechanical shaker was set to oscillate at 50 Hz. The sampling frequency is set to 5 kHz. Accelerometers outputs are shown in Fig. 2.9. Note that both the power spectrums have their peak values at 50 Hz, which comply with the dominant mechanical frequency from the shaker. The coherence of the measured signals of both ADXL001-70 and conventional accelerometer is shown in Fig. 2.10. It has the highest correlated coefficient when the frequency is close to 50 Hz. This shows that both sensors have successfully obtained the important dominant frequency information. The magnitude for harmonic frequencies between the two signals deviates when the frequency is getting larger (100 and 150 Hz) probably due to cross talks and noise contents [18].
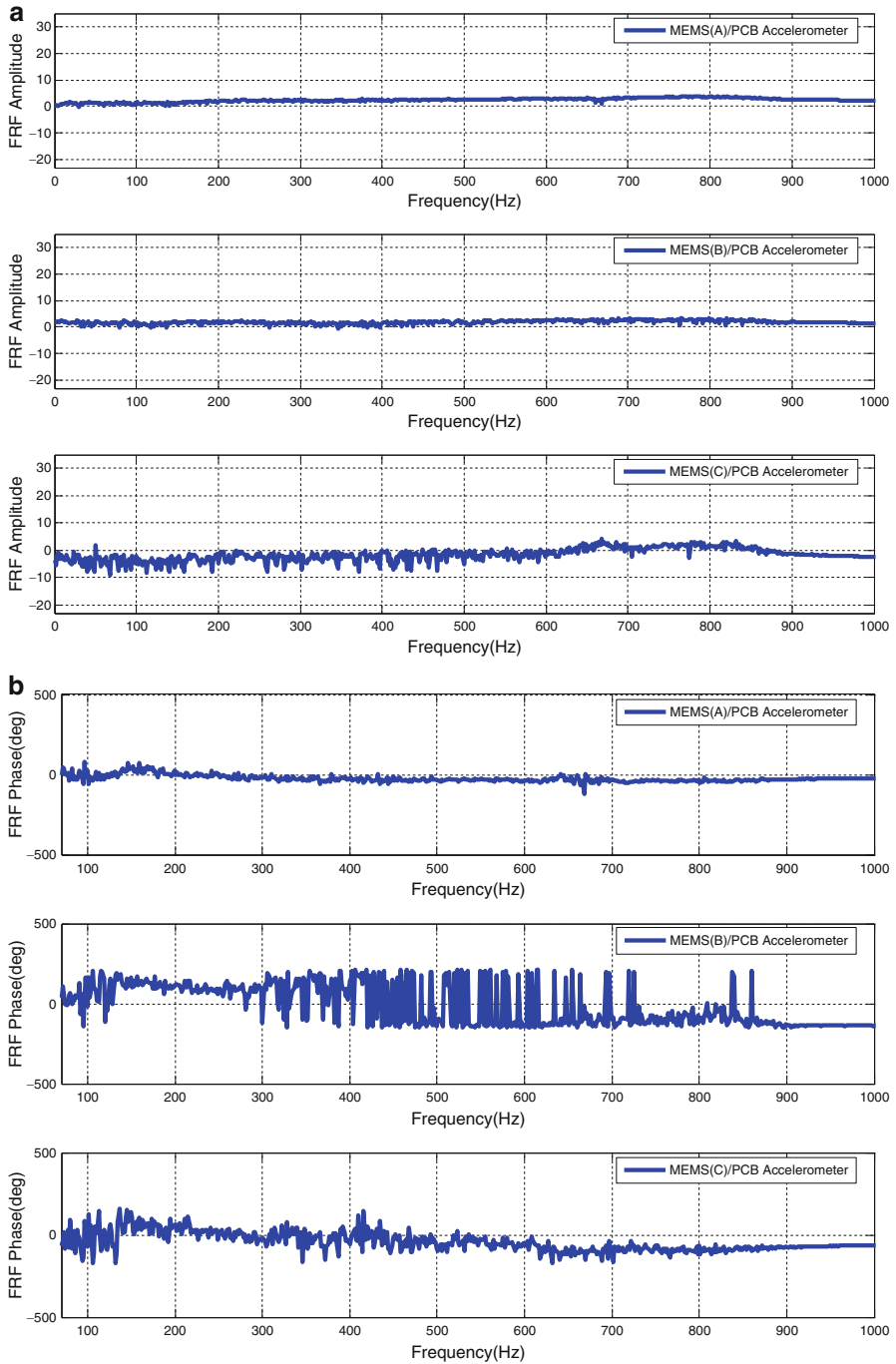
Fig. 2.6 MEMS accelerometer's response (**a**) magnitude and (**b**) phase of the frequency [4]

Fig. 2.7 Accelerometer calibrator with mounted ADXL001-70
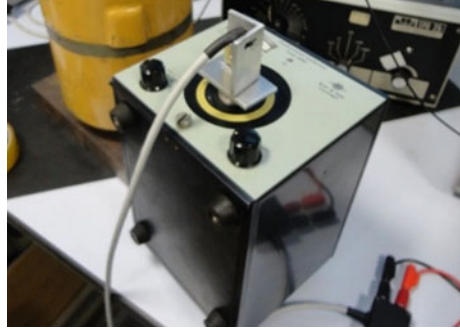
Table 2.3 ADXL001-70 calibrated result

	Parameter	
Specification	0 Hz, 0 ms^{-2} peak	79.6 Hz, 10 ms^{-2} peak
Average voltage	2.62 V	2.64 V
Peak–peak voltage	200 mV	800 mV
RMS voltage	2.66 V	2.62 V
Dominant frequency	0 Hz	80.4 Hz (1 % deviation)

4 Packaging Technique

The delicate nature of MEMS accelerometers structures necessitates the use of suitable packaging techniques to allow them to survive harsh conditions they would face when used to acquire machinery vibrations in real industrial environments. Benefits of proper packaging are to provide:

- Mechanical protection for the chip and components
- Stable temperature operating environment
- Ease of identification of sensors
- Means of mounting the sensor to equipment under test

Details of packaging typical MEMS accelerometer chip (ADXL202AE) are explained in Figs. 2.5 and 2.6. The connections (Fig. 2.11) and components (Fig. 2.12) for implementing the MEMS' chip in an analogue voltage output configuration were determined from the supplier data sheet.

5 MEMS Implementation on Machinery Diagnostics

In this section, MEMS accelerometers are used to collect vibration data from two types of rotating machinery: a DC motor and a CNC machine. These two types of machines are widely used in today's industry, and it would be of great importance
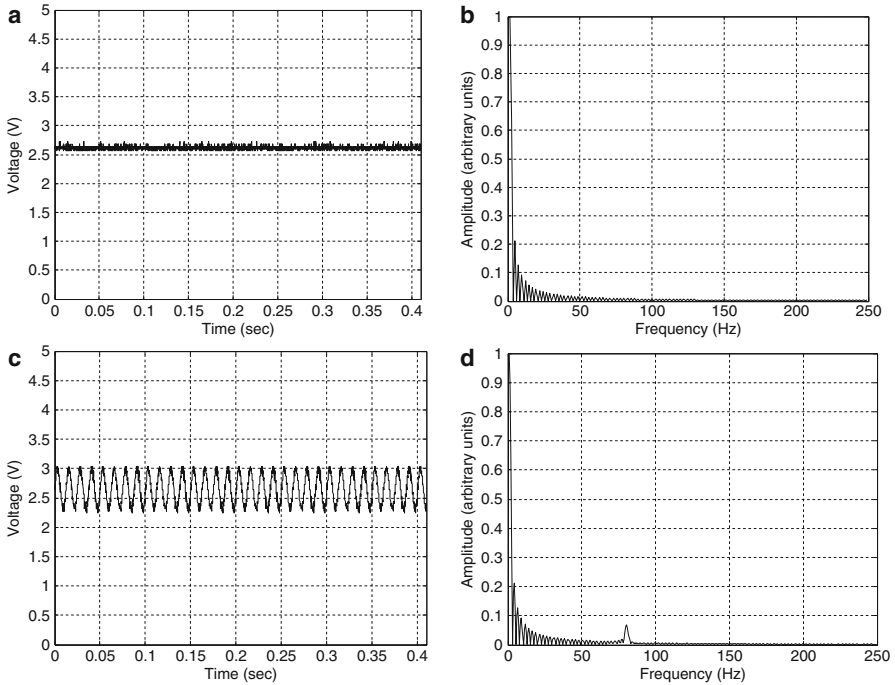
Fig. 2.8 Calibrated vibration waveform of ADXL001-70: (**a**) 0 Hz, 0 ms^{-2}, (**b**) frequency domain, (**c**) 79.6 Hz, 10 ms^{-2}, (**d**) frequency domain

if their incipient failures were successfully detected by the cost-effective MEMS accelerometers. In other words, the adopted MEMS accelerometer has to clearly show fundamental frequencies and harmonics of the monitored machines.

5.1   Vibration Measurements of DC Motor Test Rig

The DC motor test rig is shown in Fig. 2.13. It consists of a DC motor attached to a shaft supported by three roller bearing and holds a metallic disk. The motor's speed is controlled by a speed controller, which varies from 0 rpm to a maximum rotational speed of 2864 rpm. An ADXL001-70 MEMS accelerometer, placed on drive end bearing housing, is used to collect the vibration data at five different speeds: 280, 914, 1602, 2281 and 2864 rpm. Time and frequency domains of collected vibration data are shown in Fig. 2.14. The respective harmonics could be seen, in particular, at rotational speeds higher than 1602 rpm.
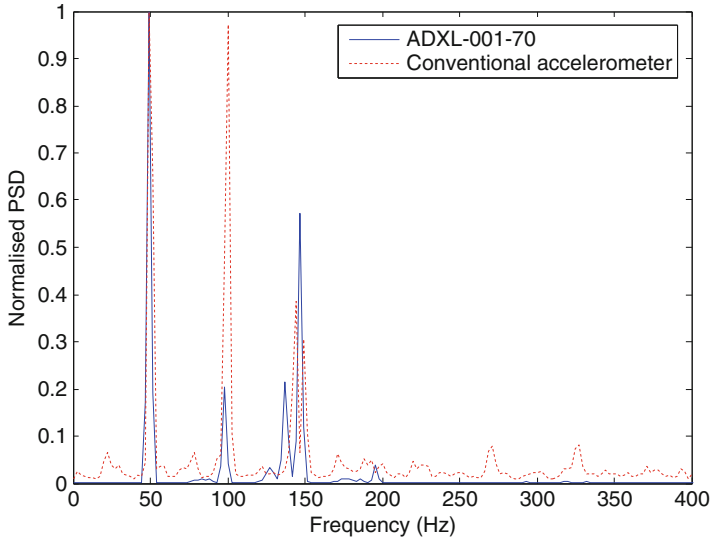
Fig. 2.9 Frequency spectrum of acquired vibration signal using ADXL001-70 and conventional accelerometer at 50 Hz
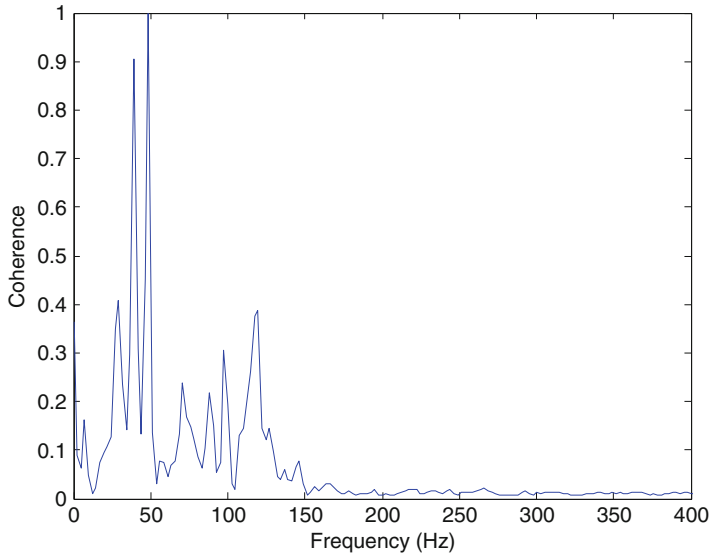
Fig. 2.10 Coherence between the vibration signals acquired using ADXL001-70 and conventional accelerometer
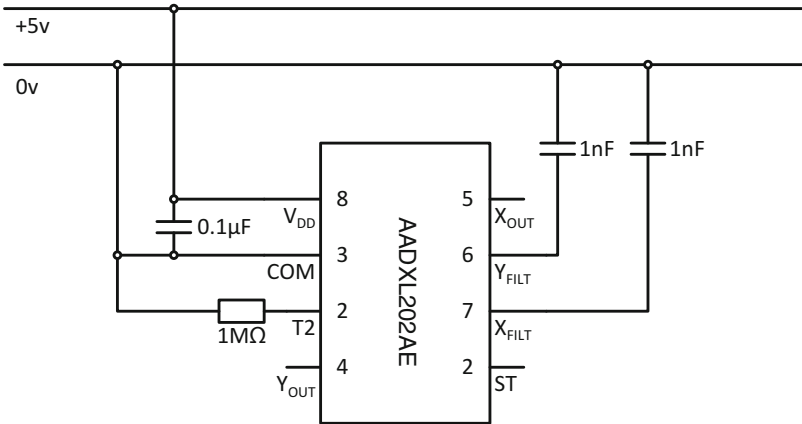
Fig. 2.11 MEMS accelerometer wiring diagram
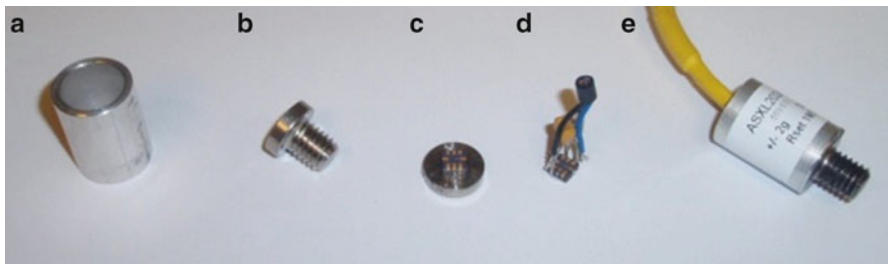
Fig. 2.12 MEMS packaged sensor components: (**a**) sensor outer casing (mild steel), (**b**) threaded mounting bottom plate, (**c**) flat mounting bottom plate with MEMS affixed, (**d**) MEMS with components attached, (**e**) Packaged MEMS sensor
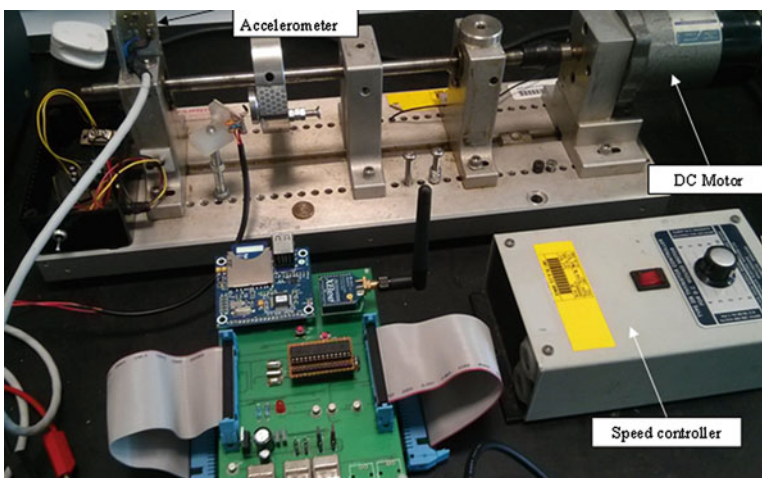
Fig. 2.13 Experimental set-up of DC motor vibration monitoring
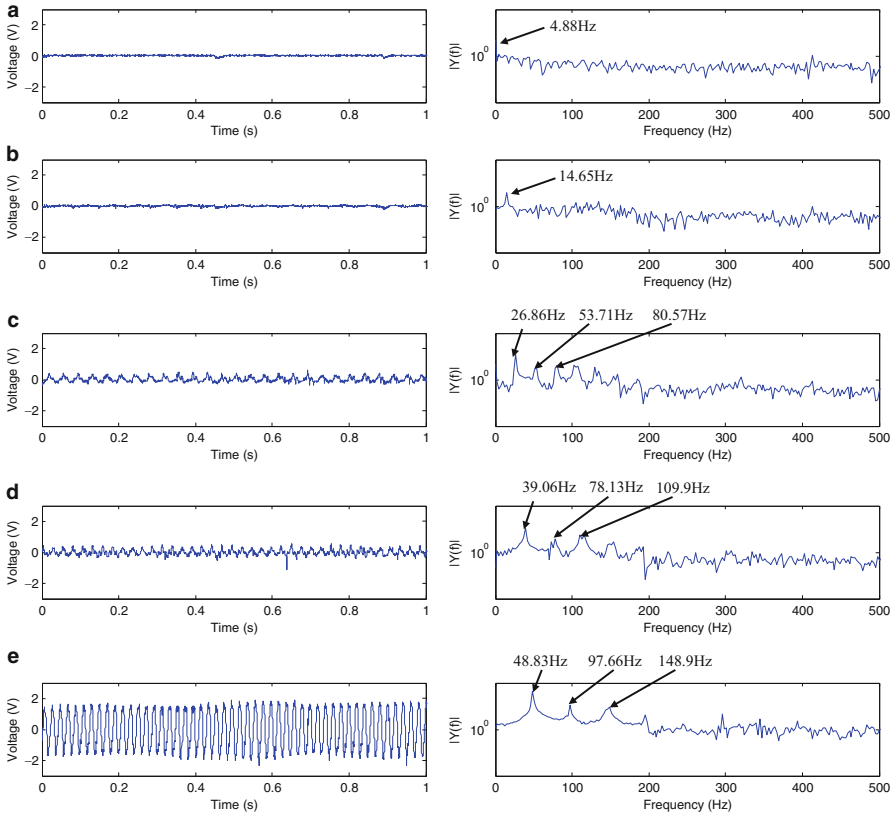
Fig. 2.14 DC motor vibration and power spectrum measured by ADXL001-70 at different rotating speeds: (**a**) 280 rpm (4.67 Hz), (**b**) 914 rpm (15.23 Hz), (**c**) 1602 rpm (26.7 Hz), (**d**) 2281 rpm (38 Hz), (**e**) 2864 rpm (47.73 Hz)

5.2 Vibration Measurements of CNC Machine

The MEMS accelerometer is placed on CNC machine in a typical industrial environment as shown in Fig. 2.15. The CNC machine operates under a speed of 2400 rpm. Respective time and frequency domains of the vibration signals are plotted in Fig. 2.16. Experimental results show that the important fundamental frequencies and their harmonics are captured by the MEMS accelerometer.

The actual vibration amplitude for the machine was about 0.2 g peak ($\pm 0.01\,g$—using hand-held commercial vibration monitoring device) which gives clear evidence on the correctness of the vibration amplitude indicated by the MEMS accelerometer.
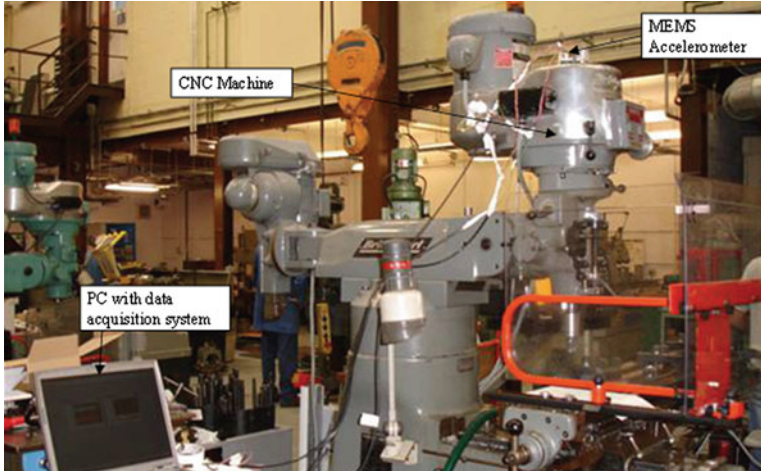
Fig. 2.15 Experimental set-up of CNC machine vibration monitoring
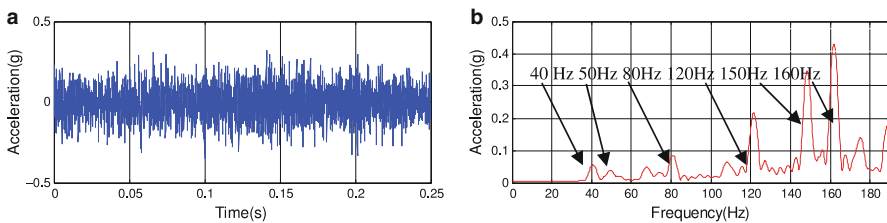
Fig. 2.16 CNC machine vibration measured by ADXL001-7 at speed of 2400 rpm (**a**) time domain and (**b**) frequency domain [4]

The CNC machine was run at 2400 rpm, and at this speed the fundamental speed and harmonics are 40, 80 and 160 Hz which were clearly presented in the frequency domain of Fig. 2.16.

6 Design and Implementation of MEMS-Based Smart Sensors

Previous sections have outlined the characteristics and fundamentals of MEMS accelerometers. MEMS accelerometers are envisioned as future candidates in realising the practical deployment of smart sensing technology. Smart sensor/transducer provides more complex functions than raw signal acquisition by integrating embedded system into the sensing architecture. Possible features of smart sensors include programmable signal conditioning, on-board signal processing and analysis, simple decision-making and remote communication. Generally, there are two major advantages of using MEMS accelerometers over their conventional counterparts for smart sensing development:

- *Low power consumption*: A conventional accelerometer requires charge amplifier to amplify the induced acceleration by conventional accelerometer. Such an amplifier usually requires a supply voltage higher than that of an embedded system (which average operational voltage is between 1.65 and 5 V and about 15 mA).
- *Low cost of deployment*: The cost of an MEMS accelerometer could be less than 10 % compared with commercially available accelerometers. The average price of MEMS has been declining since 2009, and it is believed that it will continue to do so in the next few years (predicted figure from \$1.01 to \$0.90). High performance accelerometer such as ADXL001 has an average market price at \$30, which is at least 50 % cheaper than commercially available conventional accelerometer.

To validate the feasibility of MEMS accelerometer for smart sensing application on rotating machinery diagnostics, working system architecture is designed and developed. Its respective hardware architecture could be partitioned into five main modules: sensors, embedded microcontrollers (peripheral integrated circuit and core computational unit), external data storage module and wireless communication module. The architecture design is shown in Fig. 2.17. The working prototype is also developed as shown in Fig. 2.18.

The sensing element—MEMS accelerometer—is responsible for transducing physical movement into electronic signals. The resulting electronic measurements are then sent to the peripheral computational unit, where necessary analogue signal conditioning is carried out. The processed analogue signals are then transmitted to computational core for signal processing, analysis and fault diagnostics. The quantised data is logged and interrogated in removable data storage device (universal serial bus (USB) flash drive or secure digital (SD) card) through data storage module. The end-result and diagnosis report is transmitted to another system platform through wireless communication module. A flowchart to represent the workflow of the developed smart sensor is illustrated in Fig. 2.19.
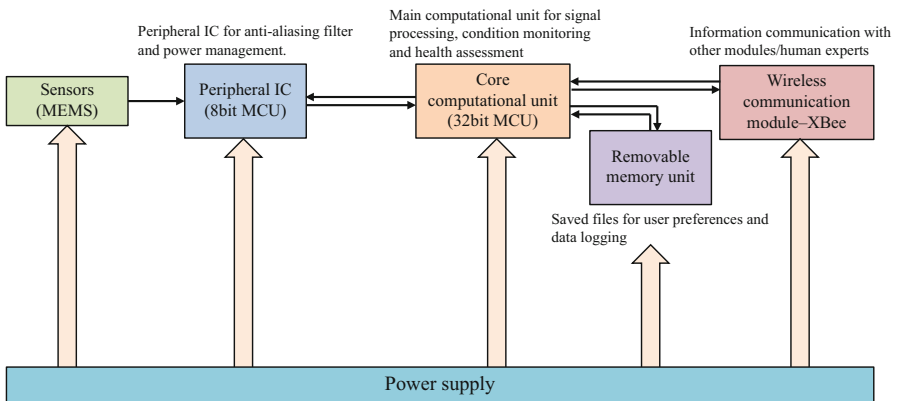
Fig. 2.17 System architecture of MEMS-based smart sensor
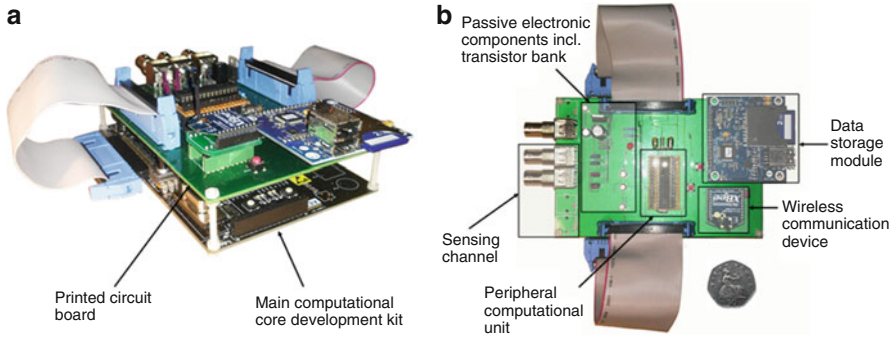
Fig. 2.18 Working prototype of MEMS-based smart sensor system (**a**) prospective view and (**b**) top layout view

On-board vibration measurement is one of the crucial elements for MEMS-based smart sensor. For example, to sample a 16-bit (2-byte) vibration signal for 3 s in 5 kHz requires at least 30 kB data memory ($2 \times 3 \times 5000 = 30$ kB). Such data load is hectic to be logged into internal memory of commercially available microcontroller, as tabulated in Table 2.4. To tackle the above design challenge, an external data storage module is necessary. The end terminal of data storage module is removable mass storage device, such as USB thumb-drive or secure digital (SD) card. This has extended high volume of data storage, which has flash memory counted in gigabyte.

The developed smart sensor has the ability to collect raw vibration, compute statistical features and send analysed result to user interface autonomously. Advanced signal processing or artificial intelligence could be performed on-board for complex machinery diagnostics. Such system has significant advantage in reducing power, cost and communication overhead in future sensing technology.

In terms of sustainability, MEMS-based smart sensor has potential to integrate energy harvesting technology in the future, which is another trending topic in modern engineering. Most commercial digital integrated circuit provide programmable sleep/hibernate mode to conserve energy consumption. Respectively, the power characteristics of developed smart sensor are tabulated in Table 2.6.

MEMS-based smart sensor integrates a wireless communication module such as XBee or XBee-PRO. Its modular design architecture and direct digital interface have enabled relatively fast transmission between the smart sensor and other sensing platform. Several customised commands are pre-programmed to access smart sensor through computer's user interface as tabulated in Table 2.5. They are coded in 8-bit ASCII character data-type, providing good readability for the user to monitor the console.

The developed smart sensor has the ability to collect raw vibration, compute statistical features and send analysed result to user interface autonomously. Advanced signal processing or artificial intelligence could be performed on-board for complex machinery diagnostics. Such system has significant advantage in reducing power, cost and communication overhead in future sensing technology.
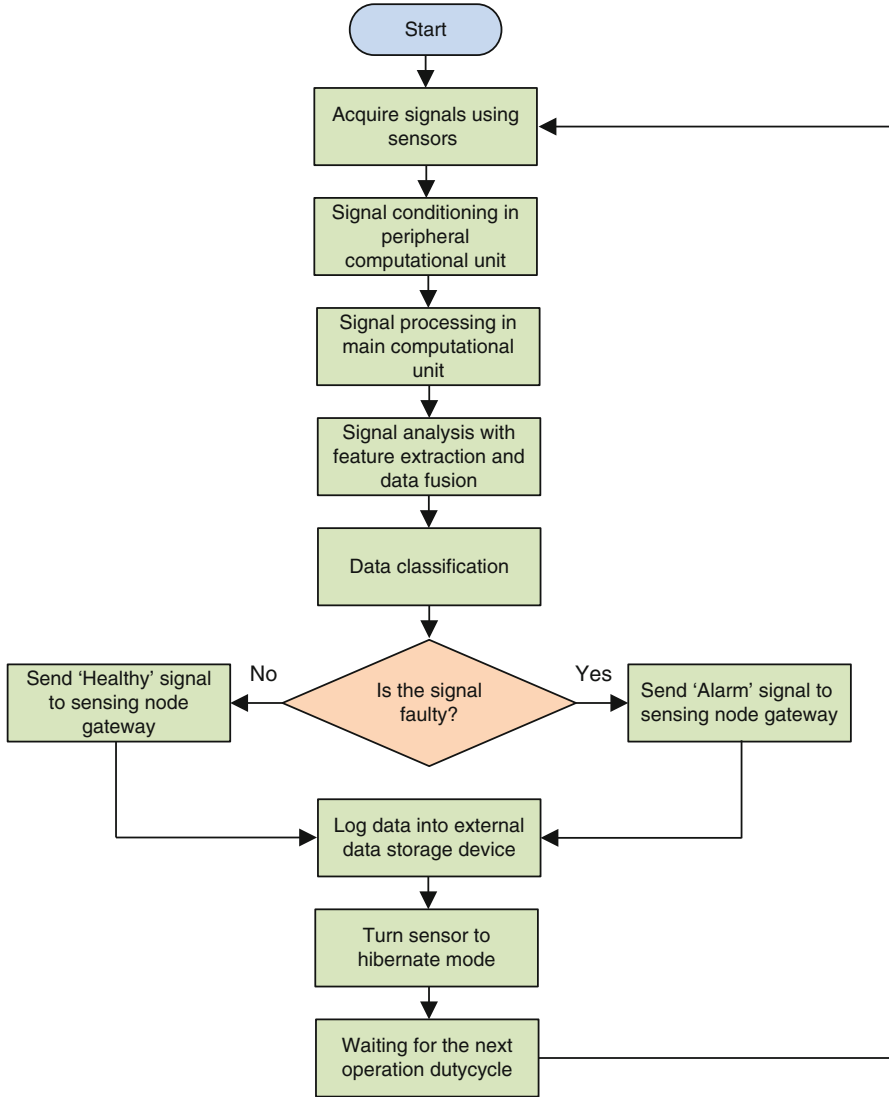
Fig. 2.19 Workflow of smart sensor

In terms of sustainability, MEMS-based smart sensor has potential to integrate energy harvesting technology in the future, which is another trending topic in modern engineering. Most commercial digital integrated circuit provide programmable sleep/hibernate mode to conserve energy consumption. Respectively, the power characteristics of developed smart sensor are tabulated in Table 2.6.

Table 2.4 Available memory capacity of various commercial microcontrollers

Microcontroller	Manufacturer	Maximum flash memory (kB)	Data memory (RAM) (kB)
PIC32MX575F512L	Microchip	512	64
dsPIC33EP512MU814	Microchip	536	52
AT32UC3C0512C	Atmel	512	64
MC56F84789	Freescale	256	32
CY8C5868AXI-LP032	Cypress	256	64

Table 2.5 Commands to communicate with IMSEM through XBee module

Commands	Description
Menu	Shows menu of the IS
Initialise	Reset or initialise the IS together with the SD card module
Read	Read data that has been stored in the SD card of IMSEM
Auto	Start monitoring process with time-driven mode
Calculate	Calculate feature components of target signals in SD card
Debug	Reserved for debugging purpose

Table 2.6 Power budget of smart sensor

Parameter	Value
Maximum rated voltage (Identified from the system modules' data sheet)	7.5 V
Rated voltage	5.6 V
Minimum voltage	4.8 V
Current drawn (Operation mode)	Average 0.22 A
Current drawn (Idle mode)	Average 0.2 A
Current drawn (Inactive mode)	8 μA
Rated power consumption (Operation mode)	1.232 W
Rated power consumption (Idle mode)	1.12 W
Rated power consumption (Inactive mode)	44.8 μW

7 Concluding Remarks

MEMS accelerometers offer excellent alternatives to conventional types of those transducers as there is no need to carry heavy charge amplifiers, but the choice has to be made according to specifications and through validation tests. MEMS sensors have also to resist harsh environments using an appropriate packaging.

There are a number of MEMS accelerometers that could be used for monitoring machinery vibrations. The technical specifications for each of them should be well understood and must be suitable for the intended application. MEMS accelerometers are of sensitive and breakable structure, and thus they should be suitably packaged

to withstand harsh environments without affecting their specifications. Sensitivity, resolutions and frequency range are the most important specifications that should be taken into account when choosing MEMS accelerometers

Adequate performance tests of MEMS accelerometers should be conducted under different excitations signals including sinusoidal, impulse and random. The measured responses of the MEMS accelerometers shall also be compared with a well-calibrated type accelerometer. Moreover, their outputs contain high amount of noise and extra un-interpretable peaks which necessitate full understanding of the measured vibration signals.

References

1. Albarbar A (2012) Vibration and noise of diesel engines. Lap Lambert Academic, Saarbrücken. ISBN 978-3-659-11039-9
2. Shi H-Y, Wang W-L, Kwok N-M, Chen S-Y (2012) Game theory for wireless sensor networks: a survey. Sensors 12(7):9055–9097
3. Ibrahim G, Albarbar A, Abouhnik A, Shnibha R (2013) Adaptive filtering based system for extracting gearbox condition feature from the measured vibrations. Measurement 46:2029–2034
4. Albarbar A, Sinha J, Starr A (2009) Performance evaluation of MEMS accelerometers. J Meas 42(5):790–795
5. Albarbar A, Mekid S, Starr A, Pietruszkiewicz R (2008) MEMS accelerometers! Are they suitable for machinery condition monitoring? Int J Sensors 8(9):784–799
6. Growing in maturity, the MEMS industry is getting its second wind Status of the MEMS Industry report, Yole Development. http://www.yole.fr//iso_upload/News/2015/PR_MEMSMarkets_YOLE_May2015.pdf. Accessed Jan 2016
7. Plaza J, Collado A, Cabruja E, Esteve J (2002) Piezoresistive accelerometers for MCM package. J Microelectromech Syst 11(6):794–801
8. Biefeld V, Buhrdorf A, Binder J (2000) Laterally driven accelerometer fabricated in single crystalline silicon. Sensors Actuators A Phys 82(1):149–154
9. Liu C, Kenny TW (2001) A high-precision, wide-bandwidth micromachined tunneling accelerometer. J Microelectromech Syst 10(3):425–433
10. Kim S-C, Wise KD (1983) Temperature sensitivity in silicon piezoresistive pressure transducers. IEEE Trans Electron Devices 30(7):802–810
11. Doll JC, Corbin EA, King WP, Pruitt BL (2011) Self-heating in piezoresistive cantilevers. Appl Phys Lett 98(22):223103
12. Zhou X, Che L, Liang S, Lin Y, Li X, Wang Y (2015) Design and fabrication of a MEMS capacitive accelerometer with fully symmetrical double-sided H-shaped beam structure. Microelectron Eng 131:51–57
13. Xie H, Fedder G (2000) CMOS z-axis capacitive accelerometer with comb-finger sensing. In: Proceedings of IEEE Micro Electro Mechanical Systems (MEMS), pp 496–501
14. Bao M (2005) Chapter 1—Introduction to {MEMS} devices. In: Bao M (ed) Analysis and design principles of MEMS devices. Elsevier Science, Amsterdam, pp 1–32
15. Analog Devices (2015) MEMS accelerometers. Analog Devices. http://www.analog.com/en/products/mems/mems-accelerometers.html. Accessed 2 Oct 2015
16. Roylance LM, Angell JB (1979) A batch-fabricated silicon accelerometer. IEEE Trans Electron devices 26(12):1911–1917
17. Shaoqun S, Jian C, Minhang B (1992) Analysis on twin-mass structure for a piezoresistive accelerometer. Sensors Actuators A Phys 34(2):101–107
18. Badri AE, Sinha JK, Albarbar A (2010) A typical filter design to improve the measured signals from MEMS accelerometer. Measurement 43(10):1425–1430

Chapter 3
Highlights in Mechatronic Design Approaches

Edwin Dertien and Stefano Stramigioli

From the 'first time right' approach to iterative design - where mechatronic design meets the maker movement

Abstract In the recent years a major change in the engineering process of mechatronics and robotics has taken place. In various design oriented laboratories around the world a shift can be recognised from a focus on analysis, simulation and modelling combined with outsourcing hardware design to the use of digital fabrication tools (laser cutter, 3D printer) allowing a cyclic (iterative) design process inside in the lab. This chapter aims to give an overview of the impact of this change, using many examples from various projects, and will share some insights and lessons learned for facilitating and implementing this process.

1 Introduction

The term digital fabrication has been introduced (or deepened) by Neil Gershenfeld [9] in his book 'Fab: The Coming Revolution on Your Desktop'. The topic (and term) are very popular describing the changes currently taking place in the world of engineering and academia. A recent publication by Bezzo et al. [2] lists myriad new ways of manufacturing robotic parts, new technologies, materials and unprecedented levels of integration. The proposed lab format by Gershenfeld has not only inspired hundreds of 'FabLabs' around the globe, but also inspired research labs worldwide to (re)incorporate manufacturing.

E. Dertien (✉) • S. Stramigioli

Robotics and Mechatronics Group, Faculty of Electrical Engineering,
Mathematics and Computer Science, University of Twente, Enschede, The Netherlands
e-mail: e.dertien@utwente.nl; s.stramigioli@utwente.nl

Besides offering new materials, technologies with a high level of integration (soft materials with embedded sensing, multi-material print in one run) the technology of digital fabrication also allowed a process optimisation with a large impact in our lab: Kaizen in Mechatronics. The term Kaizen (Japanese for 'improvement') has become famous by Toyota's successful business strategy, described by Masaaki Imai [11]. This notion has been used as basis for many iterative design processes such as SCRUM in software development. The application of this principle to mechatronic design has yielded systems in our lab with a level of integration which was hitherto unthinkable in lab prototypes.

Thanks to the concept of digital fabrication, 3D printing technology and other (2D) manufacturing techniques using laser cutter, CNC milling machines and the like are changing the way mechatronic research is being conducted. In this chapter we will discuss the changes it has brought to our lab, the changes we see worldwide, some of the high potentials and risks that come with it. For the tools to have maximum impact we will argue that the criteria *accessibility, availability and visibility* are key in successful implementation and maximising potential of this technology. As most fundamental benefit we see the possibilities of rapid iterative design cycles in mechatronics. Most impact in future will is expected from tight integration of these rapid design cycles in physical hardware with existing technology for simulation and development.

As one of the illustrative cases in this chapter the development of a pipe inspection robot for small diameter gas distribution mains is used. This project has been described in the author's thesis [5].

2   Rapid Iterative Design Cycles

Every design process (if not every scientific process) has a cyclic nature. Based on observation a hypothesis is formulated which is proven (or disproved) by designing an experiment and evaluating the results. Based on the observed outcome a new hypothesis can be formulated and the works continue.

For design and engineering normally a similar approach is taken. For a given problem a solution is designed which is analysed, tested, simulated, (and if everything looks promising) constructed, built and evaluated. The role of engineering as an academic discipline is discussed, for example, by Herbert A. Simon in his book 'The Sciences of the Artificial' [17], who describes the divide between teaching 'design' and teaching 'analysis'. The method of engineering is investigated in further depth by, among others, Schön in 'The Reflective Practitioner' [16] and Vincenti in 'What Engineers Know and How They Know It' [19].

This chapter does not intend to contribute to the discussion on the differences between engineering and academia or to present a detailed analysis of how the choice of *method* of design and engineering influences the end result. Yet the last decade a major change in engineering process has taken place. This change might be the influence of the *zeitgeist* or a deliberate choice, or both.

Since engineering hardware is expensive the realisation of a given system has to be right 'the first time', so all major effort goes to validation of the concept and implementation choices beforehand. This is a very complex process when there are many unknown variables, for instance, specific details on the environment in which a given system has to operate.

In the mechatronic approach as practiced in our lab, for example, described by Schipper [15] in 2001, a lot of effort, time and energy is devoted to the analysis and testing of a potential solution, before considering constructing a prototype for evaluation. A normal course of action for a project was to devote most of the time to simulation of a possible solution (and thereby improving the modelling and simulation tools) in order to get the expensive hardware right the first time.

This 'first time right' approach has a lot to say for it. It minimises costs of professional designers, constructors, expensive parts since the whole exercise has just to be taken once.

3 Simulation

The typical modelling and simulation tool of choice for mechatronics at our research group was, besides the ever popular Matlab, a package of a spin-off company from the group called 20sim.[1] 20sim focuses on energy based modelling of dynamical systems. It provides excellent toolboxes for 3D mechanical structures calculating inverse and forward kinematics, all using a port based approach so forces, torques, linear and rotational velocities in 3D space are modelled as twists and wrenches as described by Stramigioli and Bruyninckx [18].

A large number of projects in the past have proven the validity of this tool. One of the authors personal experiences was the design and construction of a passivity based biped robot called 'Dribbel' shown in Fig. 3.1, a 10 kg planar walker with knees [4].

An interesting problem in port based modelling is the simulation of hard contacts such as the impact of robot feet on the ground. In order to allow the simulation to work the model has to be power continuous (no discontinuities such as suddenly setting a velocity to zero or resetting the position integral are allowed). That means that for hard contacts normally a spring-damper model is introduced (referred to as the Hunt–Crossly model), its application to robotics modelling discussed recently in the work by Diolaiti et al. [8]. In this model a spring-damper combination with high damping and high spring constant, which will be coupled to the simulated moving body upon impact.

After realising the robot and matching the walking performance with the simulated outcome both seemed to match fine. Figure 3.2 shows both the walking robot and simulation. One of the outcomes of the simulation was that the simulated robot walked much better when spring constant (and especially damping) in the

---

[1]http://www.20sim.com
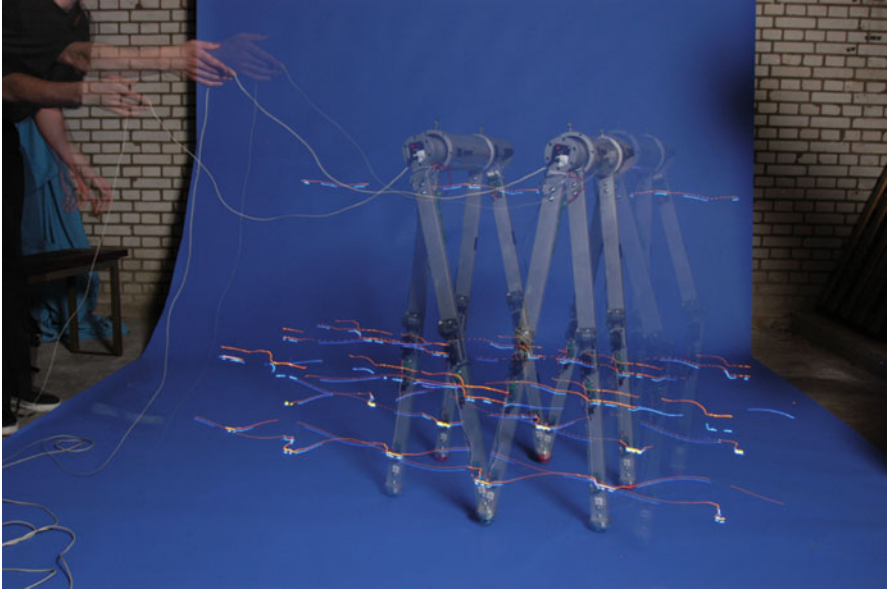
Fig. 3.1 Planar passivity based biped 'Dribbel'
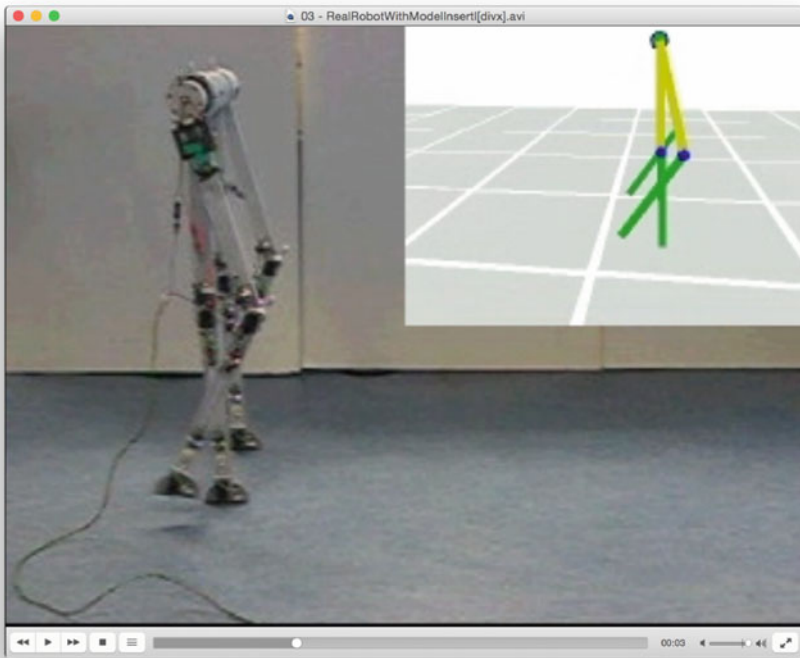
Fig. 3.2 Matched simulation of 'Dribbel' in 20sim

contact model was reduced. This makes sense in a way. Damping takes away energy from the walking motion which might be returned on impact in order to improve the walking motion and make it more energy efficient. Modern prosthetic devices using springs made famous by the 'Blade Runner'[2] verify this.

The real-world implications of this 'faulty contact model' were verified by out-fitting the robot with bouncing balls as feet (rubber with very low damping) which allowed for a smooth, very energy efficient (and rather bouncy) walk. In this project the 'first time right' approach roughly worked. The mechanism was simple enough (4DOF in planar simulation) where the most critical factor for matching the model with reality (the contact model) actually proved to be a useful design criterion.

In a next project (starting 2006) a similar approach was taken. A project was started to develop a robot for inspection of small diameter gas distribution mains as described by Pulles et al. [13], in collaboration with Dutch gas grid quality agent KIWA.[3] Again much effort was devoted to designing and evaluating a number of concepts, simulation of the most viable ones, finally selecting one of the simulated concepts for realisation. The detailed engineering of a prototype for this concept took a long time. Since none of the mechanical parts could be obtained as Common Off The Shelf (besides the chosen DC motors), every part needed detailed construction drawings. Most of the parts had to be outsourced for production. This meant a 2 month gap in the design process waiting for the mechanical parts. Electronics for control of the robot had been developed simultaneously (the whole point of a mechatronic exercise) so the time 'waiting' could be spent on continuing with software, testing electronics, etc. After completion of the mechanical prototype it still took half a year to conduct the first functional tests, and another year to solve all tiny (hardware) related issues necessary to have the robot move to its designated obstacle course which could be presented only much later at the ICRA conference in 2011 [6]. Needless to say that an improvement of the development process would be highly valued at that time.

4 Robot Makers

The project started out with a similar mechatronic design process as described with the previous case (the walking robot), but continued with a more 'maker' inspired methodology as described by Chris Anderson in his book 'Makers, the new industrial revolution' [1]. The maker movement refers in this case to the large worldwide group of DIY enthusiasts, open hardware and open software designers, hobbyists and other people that 'build' things. O'Reilly media coined the term 'makers' with the start of their magazine 'Make:' in 2005 and the organisation of gatherings of makers and builders called 'Maker Faires' in the USA.[4] Also the recent publication by Bezzo et al. [2] has been named 'Robot Makers' with the intended affiliation.

---

[2]https://en.wikipedia.org/wiki/Oscar_Pistorius

[3]http://www.kiwa.nl

[4]http://en.wikipedia.org/wiki/Maker_culture

Although strictly speaking many of the rapid prototyping technologies that are discussed in this chapter have been available for use throughout the entire span of the developed pipe inspection robot, they have actively used since 2010. This progression appears to coincide with the rise of aforesaid 'maker movement', but there might not be a strong causal connection that can be claimed. For example, in 2006 at the start of the project, a *Stratasys Dimension 3D printer*[5] was located at the University in the modelling workshop used for industrial design students. The use of this machine was effectively never considered for production or prototyping of the robot, although it could have done the job equally well as a machine that was bought much later, but places *inside* the robotics lab, making it available, accessible and visible for all projects.

5 Example Case: Pipe Inspection Robot 'PIRATE'

In this section a comparison will be made between an earlier developed prototype (shown in Fig. 3.3 and a rapid, iterative developed prototype of the same project, in this case the design of a robot for inspection of small diameter gas distribution mains. Among other things, an additional level of creativity enters the way a product is designed when allowing other technologies and also different design methodologies to enter a hitherto linear design process. At the University of Twente the Bachelor course Creative Technology tries to put different design paradigms to practice, sparking a discussion about the way chosen development methodologies influence engineering science as also, for example, discussed by Resnick et al. [14] at MIT.

The first prototype has been designed using the mechatronic design method as used at that time at collaborating project partner DEMCON, which underlying philosophy is described in the work by Schipper in his thesis [15]. Physically housed
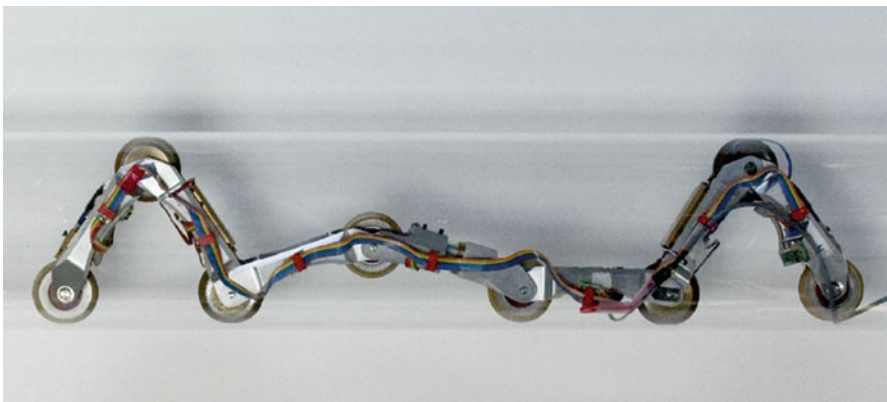
Fig. 3.3 First incarnation of a 'classically' engineered prototype of the Pipe inspection robot

---

[5]http://www.stratasys.com/
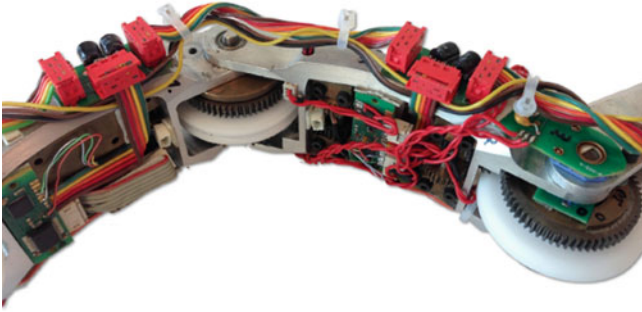
Fig. 3.4 Close up picture of the wiring of prototype 1

in the same room, the mechatronic design team consisting of a mechanical engineer, electronic engineer, software engineer and systems engineer designed the robot in a collaborative effort under supervision of two senior engineers. This initial project followed the classical engineering approach using decomposition, realisation of sub-components, integration and eventually testing and evaluation.

A modular design approach was in this case interpreted as designing separate modules with each of their specific function. This allowed the engineers to narrow their focus 'per module' instead of focussing on an overall design. In later prototypes a setup using mostly identical modules has been chosen.

The approach was mostly aimed at a 'first time right' approach, meaning that once all the mechanical design drawings had been fixed, the drawings were processed for manufacturing, the design drawings were shipped to a manufacturer and after a relative long period (8 weeks) of production and shipment a start could be made in assembling the robot.

Especially the amount of wiring necessary for the motor control boards and the inflexible installation took a lot of time and effort. Eventually the wiring had been mounted at the outside of the modules as shown in Fig. 3.4, acting as obstacles during tight manoeuvres in pipe joints. Although the placement of the wires had been discussed during the design phase, the choice had to be made to postpone this for later, since taking in the routing of cables and placement of connectors was a too large effort in this 'first time right' approach. All in all the 'first time right' approach takes a long time and lacks possibilities for quickly exploring alternatives.

6 Additive Manufacturing

During the course of the project a change took place in the world of desktop fabrication. Due to this maker movement and the FabLab[6] concept, 3D printers, laser cutters and CNC routers have made a leap from the factory to the desktop.

---

[6]http://en.wikipedia.org/wiki/Fab_lab

Accessibility, *visibility* and *availability* of production methods in a lab or at home increase the usage dramatically. This point has been made clear in the work by Mader and Dertien [12] in the context of student assignments for the bachelor track 'Creative Technology' at the UT. Similar results could be observed in different generations of students working on design assignments, having access to standard production facilities nearby (i.e. wood workshop, electronics workshop), or having access to production facilities on sight (i.e.—just—one laser cutter). Based on these observations we stated that:

Only when a tool or machine satisfies the following three criteria it will have a serious impact on the design process:

- the machine or tool needs to be *visible*
- the machine or tool needs to be *available*
- the machine or tool needs to be *accessible*.

The reverse is also true: when a machine is not visible (as in, present in the lab) the machine will not be taken into consideration. When a machine is not available (as in, constantly in use), other options will be sought. When the machine is not accessible (as in, difficult to use, steep learning curves, high threshold, no information) the machine will only be used by the happy few willing to learn and adapt.

All of these arguments might explain why in an earlier stage of the project the printer in the modelling workshop of Industrial Design was not considered: it was not visible (different building), the availability was unclear (sometimes days of work by Industrial Design students, sometimes nothing) and information on workflow, file types and necessary preparation was missing.

6.1 Design Iteration Through 3D Print

After some successful trials of the robot taking mitre-bends (shown in video accompanying [6]), it was clear that no further progress could be made with the first prototype. The main reasons were the weight of the design, lack of traction torque and mostly unreliable control electronics.

An Objet Eden 250 machine was installed in the laboratory for Robotics and Mechatronics (RaM) lab in September 2010. Based on earlier stated observations this machine was bought and placed deliberately inside the working environment of the students (and not, perhaps more conveniently, in a closed cabinet, a soundproof room or one of the workshops in the building). The machine was deliberately not planned and scheduled for production (for third parties). This means that the machine was almost always immediately available for an overnight manufacturing run of new parts. The machine accepts standard STL drawings which can be generated directly from SolidWorks. No further post-processing of the drawings is necessary, so the accessibility is high.
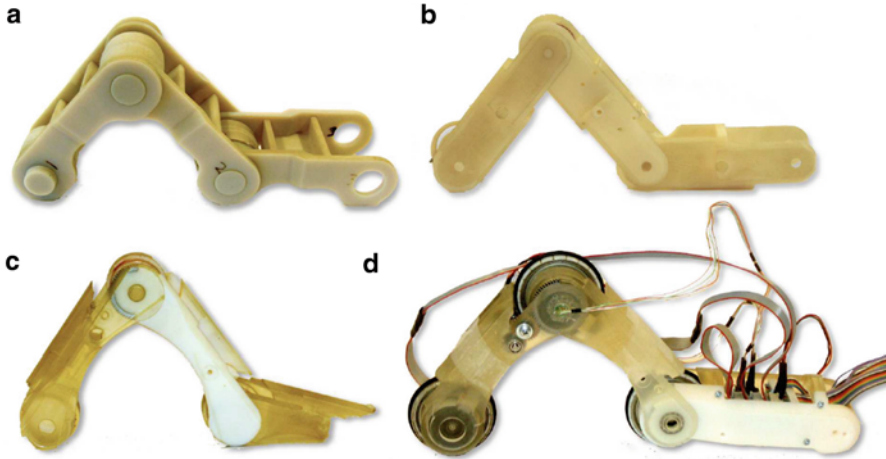
Fig. 3.5 Four iterations (in chronological order (**a,b,c,d**) of three modules of the pipe inspection robot

It took a long time before the first prototype (2006–2008) was ready for testing (first time right constraint). After manufacturing of the parts the assembly took also a long time due to small deviations in manufacturing, small errors in design, the lack of availability of necessary tooling and technicians, etc. This is in strong contrast with the process during the project by Borgerink [3] where prototypes have been produced on an almost two-weekly basis. Figure 3.5 shows three prototypes which have been subsequently designed and tested. It can be argued (and defended) that the final module design presented in this project has reached at least an equal level of complexity and completeness as the first prototype shown in Fig. 3.3.

The first two models (a) and (b) in Fig. 3.5 used printed materials for joints, the third prototype (c) and the final model (d) allowed for metal inserts such as bearings and gears. While the first prototype shown is rather bulky and not capable of moving inside even the largest of the required pipe diameters, it still yields valuable information on necessary wall thickness, placement and aligning of the drive motors, available space for the bending drives, etc. Each iteration adds more functionality and solves more design constraints. The main merit of this process is that not all design constraints have to be solved in one go, but can be tackled incrementally.

6.2 *Printed Metal Parts*

In an experiment to bring the additive manufacturing technology even further in the design process, an attempt has been done to also produce the mechanical shafts and bushes on a 3D printer. A print service nearby featured a *Concept Laser M3 Linear* machine capable of working with stainless steel. Since the tolerances of the print quality were not quite specified (everything between 0.2 and 0.6 mm deviation) in both shape and thickness, an experimental batch of the needed mechanical parts has been produced.
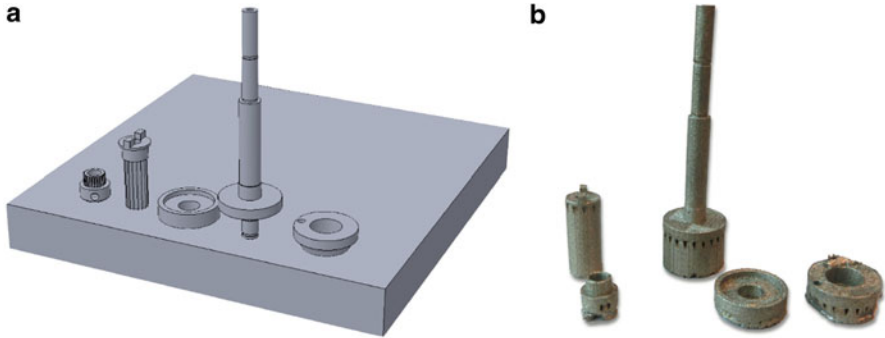
Fig. 3.6 The five experimental metal drive parts. (**a**) The parts on the build platform. (**b**) The printed batch

Fig. 3.7 Shaft with the
support material that needs to
be removed

In order to compensate for size deviations, the shafts have been printed incorporating a number of grooves on the surface—which could act as both centring aid and 'compressible' structure for an axial fit of gears and bearings. The height of the grooves has been set at $+/- 0.4\,$mm with respect to the original diameter in order to accommodate for most of the specified deviation range.

The shafts have been printed in vertical orientation (the print-bed could be much lower and the process shorter) instead of printing the shafts horizontally. Due to the support material that needs to be printed, a shaft could turn out strongly imbalanced due to the added material.

Figure 3.6a shows the parts generated in SolidWorks in the desired orientation on the build platform for print. The resulting parts are shown in Fig. 3.6b. After the printing, the support material needs to be removed. Figure 3.7 shows an extensive section of cross-hatch printed support. Figure 3.8 shows the last bit of support that is removed. After removing the support, the remaining bit needs to be sanded and polished. Unfortunately this proved to be a very time-consuming task, comparable to turning the complete parts by hand on a lathe (by a skilled technician).

The complete realised batch is shown in Fig. 3.10. The surface of the printed material remains quite rough, as can be seen in the pictures taken through a microscope (magnification set at 40x) in Figs. 3.9 and 3.10. Figure 3.11 shows a printed driveshaft with mounted Oldham coupling plate and mounted worm gear.
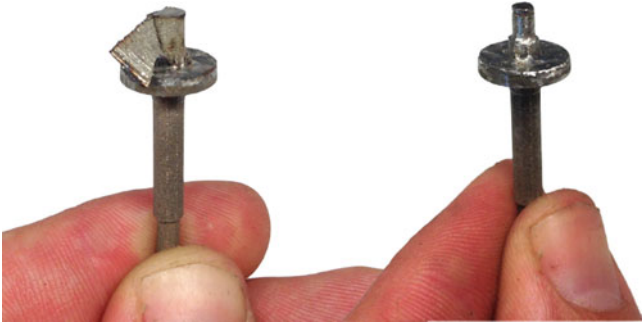
Fig. 3.8 The final stage of preparing the shaft where the metal support material is 'peeled' off layer by layer
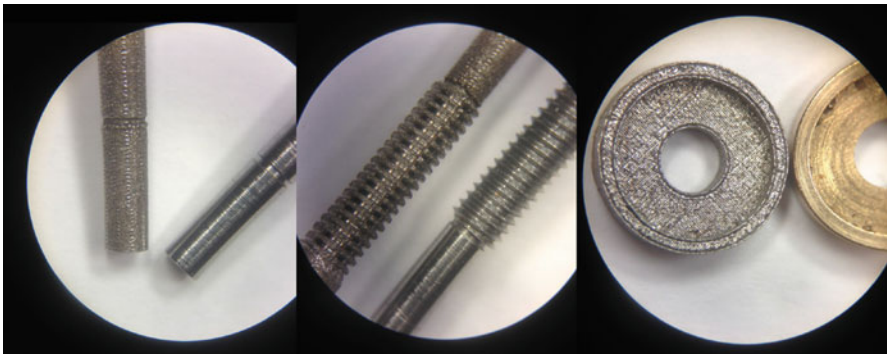
Fig. 3.9 Comparison of the printed shafts and the parts produced on (CNC) lathe

Fig. 3.10 Finished parts (after some considerable rework using a rotary tool)

Fig. 3.11 Drive shaft with fitted Oldham coupling plate and worm gear
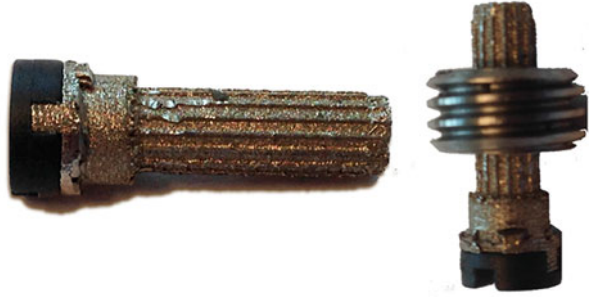
Fig. 3.12 Two drive shafts, one printed, one produced on (CNC) lathe
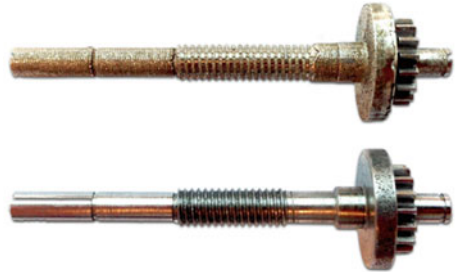
The grooves allow a reasonable tight fit although for the final assembly gluing is necessary. Also with the (CNC) produced metal parts gluing is a necessary step in the final assembly.

Figure 3.12 shows the comparison of two shafts. The gears that are fitted are stock components of HPC. Note that the thread cut in the spring shaft (the long shaft shown in Fig. 3.12) had to be cut by hand after printing. The printing quality cannot offer enough level of detail to print a reliable thread, especially not when after print some support material attached to this thread needs to be removed. To conclude, with the current printing technique the necessary level of precision for the drive parts cannot be reached. Also the price and the required time for printing and post-processing are too large to make it a part of a rapid-cycle iterative design process.

6.3 Body Material

The body structure which was originally printed in house on the *Objet Eden 250* machine has been printed for the final modules on a *FORMIGA P100 SLS* machine offered by an external print service. The main body parts are printed using PA3200: nylon reinforced with glass fibres, shown in Fig. 3.13. Since it is not easy to adapt or rework the fibre reinforced pieces, the parts which need some post-processing such as the wheels and the motor casing are printed with PA2200: 100 % nylon which can be sanded and cut reasonably well.

Fig. 3.13 Three modules printed in PA3200

7   Design for Laser Cutter

A totally different prototype for the same project using omnidirectional wheels described in [7] has been designed, constructed and tested in two weeks. As production method a 'flat' design which can be fabricated on a laser cutter has been chosen.

Following the 'digital fabrication' theme propagated by the Fablab movement, in 2010 the lab acquired a *Trotec Speedy 100* laser cutter. This machine can cut and engrave most flat materials excluding metals and other good thermal conductors and materials containing PVC. Also for designing pneumatic systems this machine has proven to be extremely versatile as shown by Groenhuis [10] in his recent work on MRI compatible robot systems.

One remarkable feature of designing 'flat' robots is that the drawing functions both as design manual and CNC file at the same time. The drawing *'is'* the design. One could select the vector drawing from the digital version of this chapter in Fig. 3.14 and send it directly to a laser cutter, resulting in almost all necessary mechanical parts for the frame (Fig. 3.15).

8   Open Micro Controller Design

During the course of the described project a different invention gradually took the world of education and design. *Arduino boards* and integrated development environment (IDE) have become a 'de facto' standard for Physical Computing and Interaction Design. Although primarily aimed at prototyping, hobbyists and 'makers', the board is also popular for rapid prototyping in mechatronic engineering disciplines. The wide variety of available extension boards, software libraries and support materials allow the board to be used as quick, standard building block.
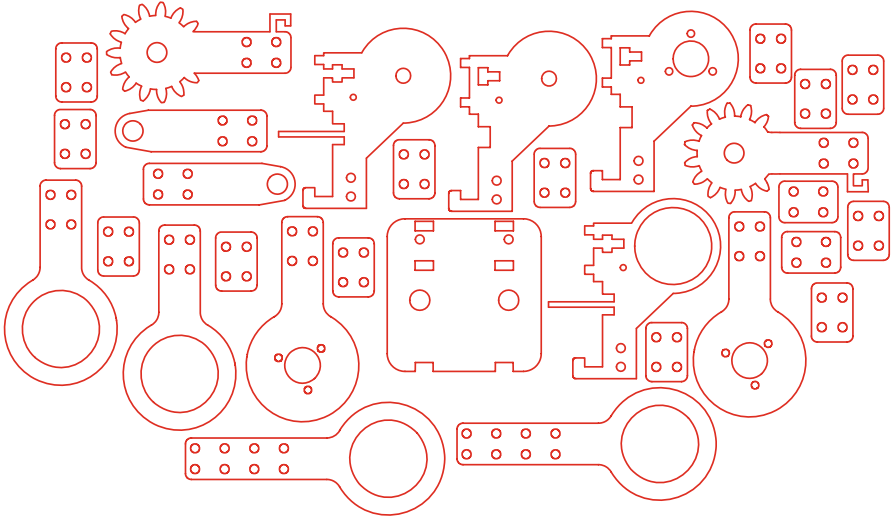
Fig. 3.14 Drawing of the parts for the omniwheel prototype
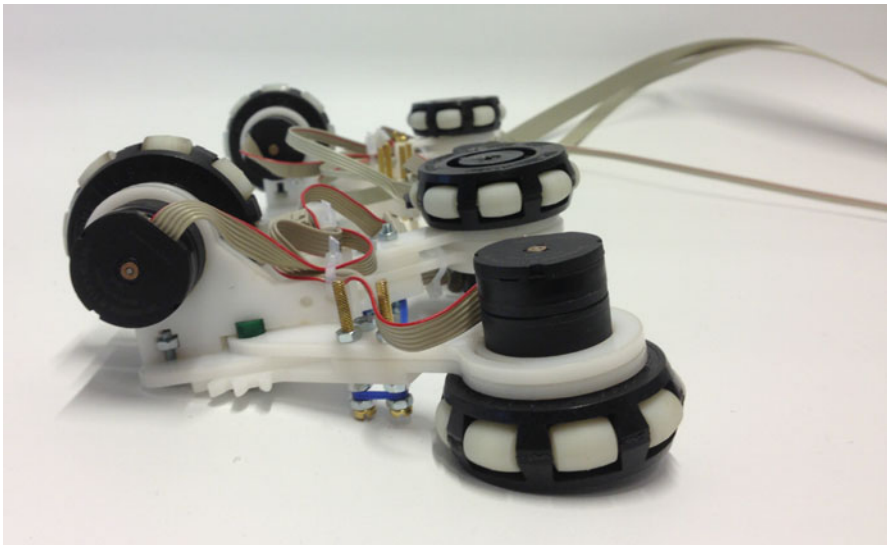
Fig. 3.15 Detailed picture of the flat construction of the omniwheel prototype

> From wikipedia[7]: *'Arduino started in 2005 as a project for students at the Interaction Design Institute Ivrea in Ivrea, Italy. At that time the students used a "BASIC Stamp" at a cost of $100, considered expensive for students. Massimo Banzi, one of the founders, taught at Ivrea. A hardware thesis was contributed for a wiring design by Colombian student Hernando Barragan. After the wiring platform was complete, researchers worked to make it lighter, less expensive, and available to the open source community. The school eventually closed down, so these researchers, one of them David Cuartielles, promoted the idea.'*

For the robot a motor control board needed to be developed, small enough to fit the minimal space available in the 3D printed housing. For the first robot model a design was made using development tools offered by the controller's manufacturer (in this case a chip and IDE made by Atmel). For the second series of prototypes the choice was made to make it fully compliant with the Arduino development environment.

The change in development process of electronics with respect to the first prototype and the subsequent boards is large. Although in principle similar hardware (AVR familie microcontrollers) and similar software (GNU GCC) have been used, the change in development process of electronics by the Arduino system is comparable to the change in development process of the mechanical system by the 3D printer. The Arduino tool set (boards, bootloaders and IDE) makes the process very accessible, and even more important: easily distributable. Many example projects are visible on the internet, libraries are readily available. This speeds up the development process dramatically.

9 Distributed Control Architecture

The developed modular hardware of the pipe inspection robot consists of many identical modules and the large number of degrees of freedom (at least 11 motors per robot) have caused something which is normally rare in prototype development: series production. The robot uses six identical drive motors with identical wheels, bearings and couplings, four identical bend motors with four identical gear sets and spring shafts. The robot uses eight identical magnetic position sensors and six identical motor control boards (not counting the motors, sensor boards and motor control boards which have to be added for an active camera module). Since wiring all these items to a central control board is simply not possible regarding the available space, also a modular approach was taken in developing the electronics, resulting in a distributed control architecture as shown in Fig. 3.16.

---

[7]http://en.wikipedia.org/wiki/Arduino
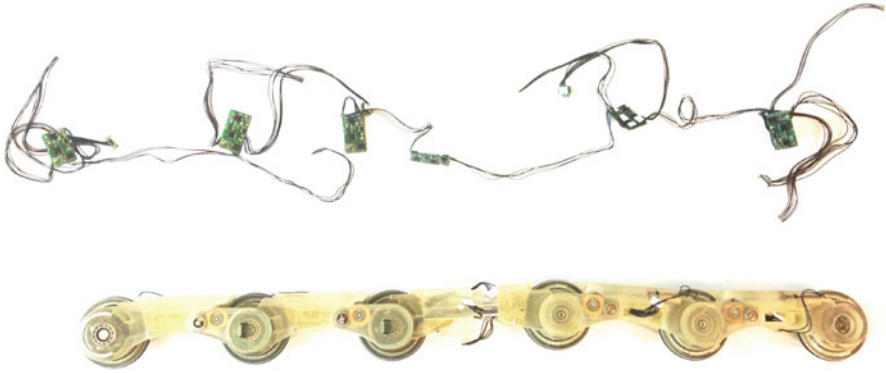
Fig. 3.16 The robot's 'nervous' system consisting of a number of control nodes
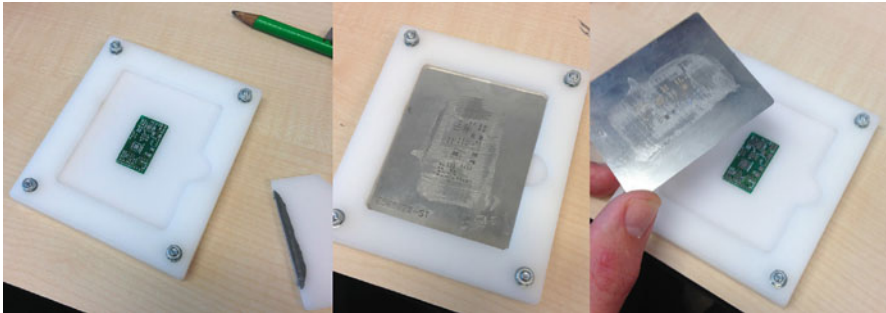
Fig. 3.17 Applying solder paste to individual PCBs

The PCBs necessary for these control nodes have been designed using the open source package KiCAD[8] and populated and reflow-soldered in house. Since the PCB pooling service produced stencils for application of solder paste, only custom tool to fixate and outline the stencil with respect to the PCB was needed to do precision application. Again, the tool was produced on the laser cutter (shown in Fig. 3.17). Also the reflow oven shown in Fig. 3.18 which has been used was an Arduino controlled converted toaster oven.

Figure 3.19 shows the control board and the sensor board which have been developed using only open source tools for development. The relative 'ease' of production, the small development time, the simplicity with which these designs can be designed, deployed and distributed are striking, when comparing it to the (relative) slow pace of development only a short time before.

---

[8]http://www.kicad-pcb.org
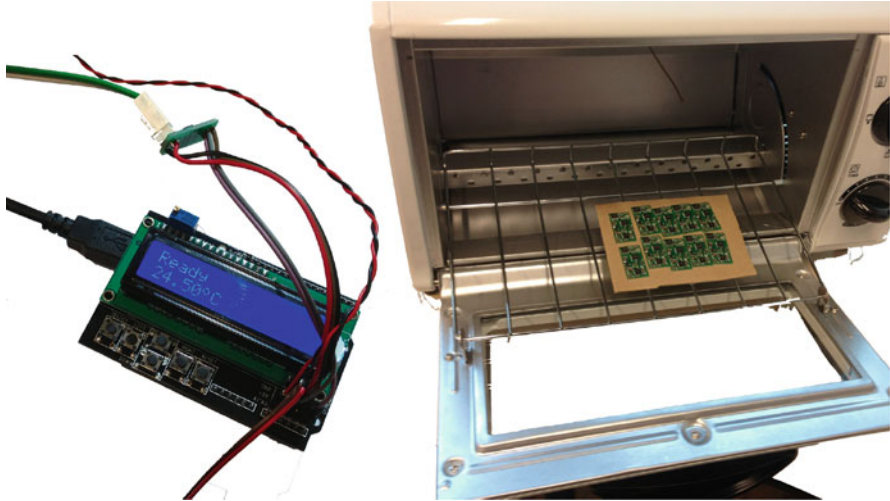
Fig. 3.18 Converted toaster oven used for reflow soldering
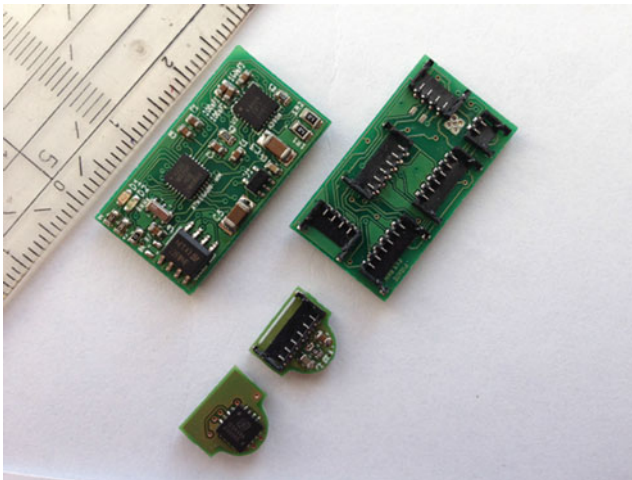
Fig. 3.19 Motor drive board and the position sensor board

10 Conclusion

In this chapter a wide number of techniques have been shown which are in itself not quite new although are gradually taking their place in the design process of mechatronic systems, allowing rapid development cycles. Visibility, availability and accessibility are key ingredients in facilitating this upheaval in usage. It is interesting to realise how the tools shape the design process (and not only vice versa, i.e. that the design process can dictate the choice in tools).

Additive manufacturing (3D printing) has proved a technology with many benefits for the design process of mechatronics. The yielded design, however, is not immediately suitable for other production methods (CNC milling, injection moulding) since the printing process allows much design freedom than conventional techniques. Fortunately this might not be the problem it seems, since the available materials for 3D printers have seen a tremendous increase, as well as providers and services offering the tooling.

One of the major contributions to the *accessibility* of the tools listed in this chapter is the support by a large community of users, closely linked to various movements such as open hardware, open source, FabLabs, MakerSpaces, HackerSpaces and other communities.

One of the major contributions to the *visibility* of the tools listed in this chapter is giving them a place directly in the space (the lab) where the design process takes place.

One of the major contributions to the *availability* of the tools is that they have not been installed to give a return in investment (money wise) regarding production costs, but that the machines are simply there, waiting for students, engineers and researchers to fill and fuel them with their ideas.

Many of the robot projects which are currently carried out are on the boundary between a laboratory prototype and an 'industrialised' robot for real use. The complexity of many of the proposed designs requires production and manufacturing on an almost industrial scale, since in many situations only a complete robot can act as 'proof of principle'.

Fortunately the tools for digital (desktop scale) fabrication which have been popularised through the 'maker' world in recent years allow 'industrial' manufacturing within the lab environment, and even go a step further by offering an unprecedented level of flexibility and agility in the design process.

References

1. Anderson C (2012) Makers, the new industrial revolution. Crown Business, New York
2. Bezzo N, Mehta A, Onal C, Tolley M (2015) Robot makers: the future of digital rapid design and fabrication of robots. IEEE Robot Autom Mag 22(4):27–36
3. Borgerink DJ (2012) Development of the second pirate prototype using rapid prototyping. Master's thesis, Control Laboratory, Drienerlolaan 5, Enschede
4. Dertien E (2006) System specifications for pirate. Technical report, University of Twente
5. Dertien E (2014) Design of an inspection robot for small diameter gas distribution mains. Ph.D. thesis, University of Twente, Enschede
6. Dertien E, Stramigioli S, Pulles K (2011) Development of an inspection robot for small diameter gas distribution mains. In: 2011 IEEE international conference on robotics and automation (ICRA), pp 5044–5049
7. Dertien E, Mozzafari Foumashi M, Pulles K, Stramigioli S (2014) Design of a robot for in-pipe inspection using omnidirectional wheels and active stabilisation. In: IEEE international conference on robotica and automation (ICRA) 2014
8. Diolaiti N, Melchiorri C, Stramigioli S (2005) Contact impedance estimation for robotic systems. IEEE Trans Robot 21(5):925–935

9. Gershenfeld NA (2005) Fab: the coming revolution on your desktop–from personal computers to personal fabrication. Basic Books, New York
10. Groenhuis V, Stramigioli S (2015) Laser-cutting pneumatics. IEEE/ASME Trans (99):1–1
11. Imai M (1986) Kaizen: the key to Japan's competitive success. Random House, New York
12. Mader A, Dertien E (2014) How to educate for creativity in creative technology. In: Arthur E, Wouter E (eds) The 16th international conference on engineering and product design education (E&PDE2014), Sept 2014
13. Pulles C, Dertien E, van de Pol HJ, Nispeling R (2008) Pirate, the development of an autonomous gas distribution system inspection robot. In: International Gas Union research Conference, IGRC 2008, Paris, France, USA, October 2008, Curran Associates
14. Resnick M, Rosenbaum E (2013) Designing for tinkerability. In: Design, make, play: growing the next design, make, play: growing the next generation of stem innovators. Taylor and Francis, New York, pp 163–180
15. Schipper D (2001) Mobile autonomous robot Twente. Ph.D. thesis, University of Twente
16. Schön DA (1982) The reflective practitioner - how professionals think in action. Basic Books, New York
17. Simon HA (1996) The sciences of the artificial. The MIT Press, Cambridge
18. Stramigioli S, Bruyninckx H (2001) Geometry of dynamic and higher-order kinematic screws. In: IEEE International Conference on Robotics and Automation, 2001. Proceedings 2001 ICRA., vol. 4, pp. 3344–3349, 2001
19. Vincenti WG (1990) What Engineers know and how they know it. Johns Hopkins, Baltimore

Chapter 4
Microrobots for Active Object Manipulation

Roel S. Pieters, Hsi-Wen Tung, and Bradley J. Nelson

Abstract Active manipulation of objects that are smaller than 1 mm in size finds its application in tasks such as assembly and pick-and-placement. Here, we present the design of a family of microrobots capable of object manipulation in a fluidic environment. The microrobots are fabricated from polymer (SU-8) with internal soft-magnetic posts (CoNi) that align to an external magnetic field and have a maximum dimension of $50 \times 200 \times 600 \, \mu m$. Actuation of the device can be enforced with either a rotating or stepping magnetic field and corresponds to the method of object manipulation. In particular, a rotating magnetic field enables a fluidic-based noncontact manipulation technique, while a stepping magnetic field enables a contact manipulation technique. The capabilities of these designs are analysed and demonstrated with respect to the generated motion and the manipulation of objects.

Keywords Microrobot • Object manipulation • Magnetic actuation • Rolling actuation • Stepping actuation • Magnetic torque • Fluidic trapping • Fluidic drag • Fabrication • MEMS • Low Reynolds number • Microrobot design • Micromanipulation

1   Introduction

The manipulation of objects or components at the micrometer scale is a challenging task. As objects become smaller, disturbances come into play that can usually be ignored for the manipulation of objects with a larger size. Examples include effects due to surface interaction (e.g. Van der Waals force and electrostatics [1]) and the dominance of viscous forces when the surrounding environment is fluidic (e.g. water [2, 3]). Most commonly, locomotion of such untethered microdevices is either executed on a planar surface or in full 3D. The design (i.e. size, shape and actuation mechanism) of the device therefore prescribes its motion capabilities and the disturbances that will act upon it. When the size of the end-effector has

R.S. Pieters (✉) • H.-W. Tung • B.J. Nelson
Institute of Robotics and Intelligent Systems, ETH Zurich, Tannenstrasse 3,
Zurich 8092, Switzerland
e-mail: rpieters@ethz.ch; bnelson@ethz.ch

the same order of magnitude as the size of the object, in general, manipulation becomes easier. Considering the microrobot as an end-effector, object manipulation differs greatly compared to manipulation with larger tools such as robotic tweezers or probes. Such tweezers are usually inserted from a certain angle and are limited in dexterity and degrees of freedom for manipulation. Moreover, when the size of the microrobot decreases, so does the force necessary for actuation. For instance, the drag force acting on a sphere in viscous flow scales with R, the radius of the sphere, while the gravitational force scales with the volume of the sphere, and thus with R^3. Additionally, when the object scales down such that the Reynolds number $Re \ll 1$, the motion of the microrobot and of objects will be more damped, and local fluidic effects can be exploited. In particular this includes the creation of a vortex due to the local rotation of the microrobot and a fluid's no-slip condition on the surface. Both can be used to aid in the lifting and/or transportation of objects. All these characteristics enable actuation techniques unconventional for macro-scale robotics and include magnetic, optical and fluidic techniques.

1.1 Related Work

Many different tools, designs and actuation techniques exist for micro- and nanomanipulation (see, e.g., [4]). A comparison regarding magnetically actuated microrobots can be found in [5, 6]. Moreover, a review on methods in automated planning and control for micromanipulation, and contact micromanipulation in particular can be found in [7, 8], respectively. Propulsion mechanisms with magnetic fields can be classified by the way energy is transferred to the robot, either force based with magnetic field gradients or torque based with alignment to a magnetic field. Examples of magnetic gradient control can be found in [5, 9, 10], while torque-based actuation techniques can be found in [3, 11, 12].

Regarding the manipulation of objects, research and applications in the microscale are single cell studies [13], minimally invasive surgery [14] and micro-assembly [15]. Other work considering manipulation using an untethered microrobot can be found in [12, 16]. Automated manipulation via fluidic trapping (i.e. pick-and-place) of beads with a similar microrobot as presented in this work is demonstrated in [17]. Additional work considering similar automated fluidic trapping strategies can be found in [16].

1.2 Micromanipulation Environment

Microrobot and micro-object manipulation is often executed in a low Reynolds number fluid environment. In this regime viscous forces are dominant and have to be overcome to achieve motion. Particular details of fluid mechanics at low Reynolds number can be found in many textbooks [18]. Other work, describing

life at low Reynolds number and robotics at the microscale can be found in [1, 2], respectively. One important effect that a low Reynolds number fluid has on bodies is the considerable influence of drag. When in motion, a translating and rotating body will be disturbed by a drag force and drag torque, respectively. This drag force and torque increases with increasing linear and rotational velocity of the body and with the viscosity of the surrounding medium. Other effects that occur due to the size scale are surface interaction forces such as adhesion or Van der Waals forces [19].

One additional property of a low Reynolds number flow that is worth mentioning is its reversibility. When $Re \ll 1$, meaning a laminar and reversible flow, the microrobot can move forwards and backwards and have an almost instantaneous flow reversal. This means that when an object is transported without contact in the vortex, a sudden motion reversal of the robot will not result in the loss of the object.

In this chapter we present a family of microrobots that are actuated via torque with a rotating and stepping magnetic field. Additionally, we show the object manipulation capabilities of these microrobot designs. The rotating magnetic field enables a fluidic-based noncontact manipulation technique, while a stepping magnetic field enables a contact manipulation technique. We emphasize that these different robotic designs result in different actuation techniques and object manipulation approaches, while the fabrication process remains the same. By altering the internal and external configuration of the microrobot, that is, by changing the size, shape and configuration of the internal soft-magnetic posts with respect to the microrobot's body, entirely different propulsion and manipulation mechanisms are possible.

2 Microrobot Design

The microrobot (see Fig. 4.1 and Table 4.1) is a wireless mobile device and has a width of 60–200 μm and a length of 150–600 μm. It is designed as a cuboid (SU-8) with internal soft-magnetic posts consisting of a cobalt–nickel (CoNi) alloy that aligns with an external magnetic field. When the field direction changes, the microrobot realigns with it [20].
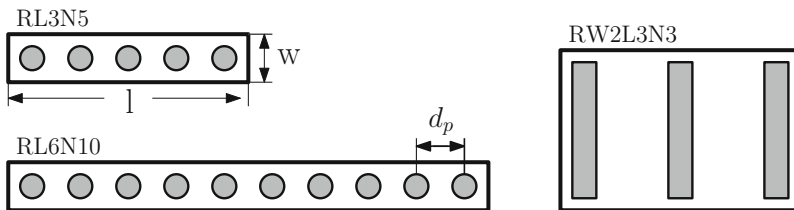
Fig. 4.1 Top view of the microrobot design. The geometry of the microrobot is defined by the length l and width w of the body. The configuration of the soft-magnetic posts differs in pitch between the posts d_p and the number of posts n_p. All cylindrical soft-magnetic posts have a diameter $D = 30\,\mu$m and a height of $h = 50\,\mu$m. The *grey area* depicts CoNi, the white area depicts SU-8. The type description of the device is explained in Table 4.1

Table 4.1 Microrobot design configuration

Name	Rod length l μm	Rod width w μm	Post pitch d_p μm	Number of posts n_p
RL3N5	300	60	$2D$	5
RL6N10	600	60	$2D$	10
RW2L3N3	300	200	$4D$	3

2.1 Design and Capabilities

To give the microrobot as much torque as possible for actuation different configurations were designed. These designs differ from each other in geometry, number and shape of soft-magnetic posts and their configuration in the device. Geometrically, these microrobots were fabricated with different length l and width w of the body. The configuration of the soft-magnetic posts varies in pitch between the posts d_p, the number of posts n_p as well as the shape (cylindrical or rectangular). All cylindrical soft-magnetic posts have a diameter $D = 30$ μm and a height of $h = 50$ μm. Several configurations are detailed in Fig. 4.1 and their type name is explained in Table 4.1. For example, a microrobot with length $l = 300$ μm, width $w = 60$ μm, number of posts $n_p = 5$, spaced $d_p = 2D$ apart, would be denoted as RL3N5.

According to [17], the torque can be increased with either a stronger magnetic field **H**, a more pronounced easy axis, or with increasing volume V of the magnetic material. The limitation of increasing magnetic material is related to the preferred magnetization direction (easy axis) of the device. In order to achieve the desired rolling motion (left two designs in Fig. 4.1) the microrobot has to be kept as symmetrical as possible along its long axis and has a transverse magnetization. This is achieved with the internal soft-magnetic posts. When the pitch between the soft-magnetic posts d_p is too small, the magnetization direction can change. The ideal configuration, therefore, depends on d_p, and defines the number of posts n_p. For increasing pitch distance d_p the soft-magnetic posts do not influence each other and behave like individual cylinders. The minimal configuration was experimentally found with $d_p = 2D$. The dimensions of the soft-magnetic posts are therefore preferred with a diameter of $D = 30$ μm and a height of $h = 50$ μm, thus having a height to diameter ratio of $h/D \approx 1.67$. This leads to a volume fraction of CoNi in the device up to 20 % (nearly 65 % by weight) allowing a high magnetization and generating sufficient torque to drive the agent in liquid environments, including in higher viscosity fluids.

Regarding the stepping actuation technique of the device (right design in Fig. 4.1), the microrobot is designed as a rectangular cuboid, with internal rectangular soft-magnetic posts. This leads to a height to diameter ratio of $h/D \approx 2.92$ offering sufficient torque to enable the stepping actuation technique. For both designs, motion is induced due to the friction of the microrobot with the supporting surface, which is explained in more detail in Sect. 3.

2.2 Fabrication

For the fabrication of the microrobot, a silicon wafer covered with a silicon dioxide (SiO$_2$) sacrificial layer is used as the substrate, followed by an evaporated titanium (Ti) and gold (Au) bilayer that is patterned by a lift-off process. A thin layer of adhesion promoter, Omnicoat, and a 50 µm thick layer of SU-8 are spincoated on the wafer and patterned into tethered structures with holes within the bodies. The holes are then filled with CoNi alloy by an electroplating process, followed by another thin layer of electroplated gold to protect the CoNi alloy during the subsequent etching step. The wafer is then diced into small chips and immersed in a buffered hydrofluoric (BHF) acid solution to etch away SiO$_2$ and release the structures from the substrate. Finally, a micro-laser milling machine is utilized to cut the microrobots from their tethers. For more details on the fabrication process, we refer to [21]. This fabrication procedure lends itself perfectly for mass-production of the microrobot and in one batch (wafer) up to 40,000 devices can be fabricated.

Due to its exceptional magnetic properties, electrodeposited soft-magnetic cobalt-nickel (CoNi) is used. The content ratio of cobalt and nickel specifies the magnetic properties of the material. This leads to tunable magnetic properties, ranging from semi-hard to very soft [22]. The used CoNi alloy contains approximately 50 % Co and has high magnetic permeability, high magnetic saturation and low remanence [a saturation magnetization of 0.85 T, and a coercivity of 20 Oe (0.002 T)]. The high magnetic permeability allows the magnetic material to generate higher force or torque in a stronger external magnetic field. A low remanence reduces the residual magnetic force when the magnetic field is turned OFF?

3   Magnetic Actuation

Actuation with magnetic fields is a convenient method for applying power to a microrobot and control its motion. Considering a magnetic field **H** that is applied in the workspace, the magnetic torque and force that is acting on the microrobot can be defined as

$$\boldsymbol{\tau} = V\mathbf{M} \times \mathbf{B} \tag{4.1}$$

where **M** is the volume magnetization in $[A/m]$ of the object of volume V, the magnetic field has a flux density $\mathbf{B} = \mu_0\mathbf{H}$ and μ_0 is the permeability of vacuum. The force acting on the microrobot is described as

$$\mathbf{F} = V(\mathbf{M} \cdot \nabla)\mathbf{B}. \tag{4.2}$$

For the designs presented in this work, the device is actuated by magnetic torque.
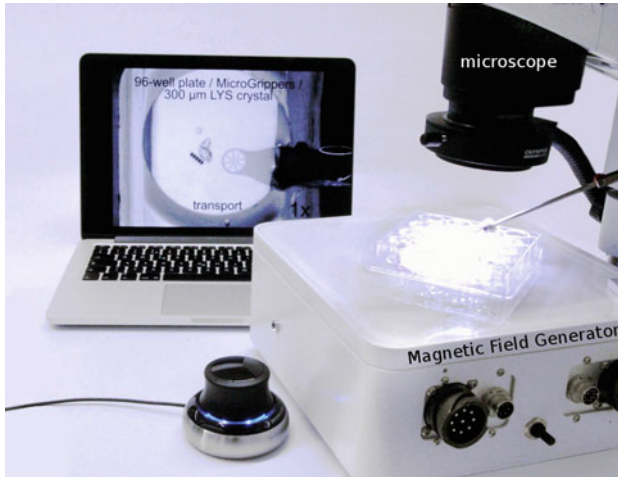
Fig. 4.2 The experimental system contains the eight-coil magnetic field generator (MFG, white box), a microscope (Olympus IX-81) and a laptop computer for user interaction. Image courtesy of MagnebotiX AG

3.1   Experimental System

The magnetic field generator (MFG) has eight electromagnetic coils as described in [23] and is commercialized by MagnetbotiX AG[1]. The system is controlled by a single computer and is capable of 5-DOF wireless control of micro- and nano-structures (3-DOF position, 2-DOF pointing orientation) within a spherical workspace with a diameter of approximately 10 mm. This allows for magnetic fields and field gradients up to 20 mT and 2 T/m. The MFG is positioned under a microscope (Olympus IX-81) to observe a container enclosing the microrobot (see Fig. 4.2). For all experiments the magnitude of the magnetic field is set to 10 mT and, as the motion of all presented devices is torque induced, no magnetic field gradient is applied.

3.2   Rotational Magnetic Fields

Figure 4.3 describes the motion of a rolling microrobot actuated with a rotating magnetic field. If the microrobot rolls without slipping, the maximum forward velocity can be determined as $\hat{v}_{\text{com},r} = 2 \cdot (w + h) \cdot \omega$, with ω the rotational frequency of the magnetic field. This assumes that the robot follows the field exactly. The forward body length of the microrobot is $w = 60\,\mu\text{m}$, resulting in

---

[1]http://www.magnebotix.com
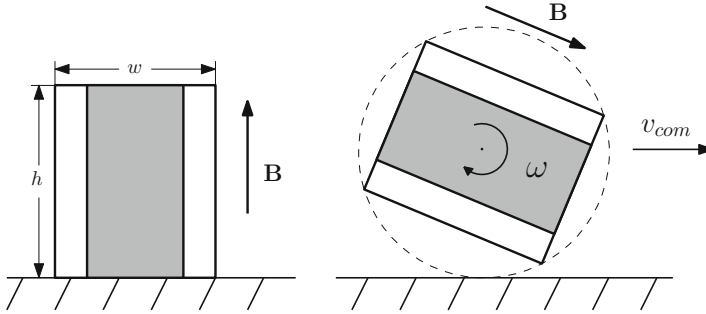
Fig. 4.3 Side view of a microrobot capable of rolling motion. Applying a rotational magnetic field **B** enables the microrobot to roll on a supporting surface with a rotational velocity ω and results in a forward motion v_{com}
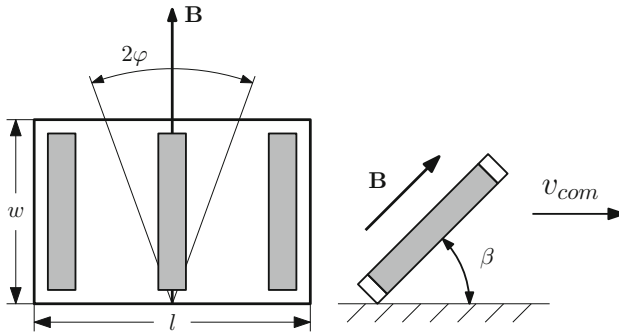
Fig. 4.4 Top view (*left*) and side view (*right*) of a microrobot capable of stepping motion. Applying a stepping magnetic field **B** enables the microrobot to tilt and slide on a supporting surface and results in a forward motion v_{com}. φ represents the stepping angle and β represents the tilt angle of the magnetic field

a forward velocity of around 16 body lengths/s at a rotation frequency of 5 Hz. The deviation of the experimental data from this ideal forward velocity indicates slip of the microrobot, as will be explained in Sect. 3.4.

3.3 Stepping Magnetic Fields

Figures 4.4 and 4.5 show the motion of a microrobot of type RW2L3N3 actuated with a stepping magnetic field. If the microrobot steps without slipping, the maximum forward velocity can be determined as $\hat{v}_{com,s} = l\sin(\phi) \cdot \eta$, with η the stepping frequency of the magnetic field. The forward body length of the microrobot is $w = 200\,\mu m$, resulting in a forward velocity of around 2.5 body lengths/s at a stepping frequency of 5 Hz. The forward velocity does not depend on the tilt angle
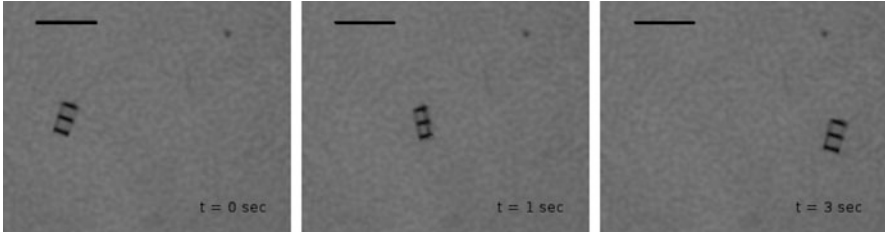
Fig. 4.5 Snapshots of microrobot actuation via a stepping magnetic field. The microrobot of type RW2L3N3 ($l = 300\,\mu\text{m}$, $w = 200\,\mu\text{m}$) moves from left to right and covers a distance of approximately 1.6 mm in the images sequence. This results in a forward speed of around 2.5 body lengths/s. The black scale bar is $500\,\mu\text{m}$ long
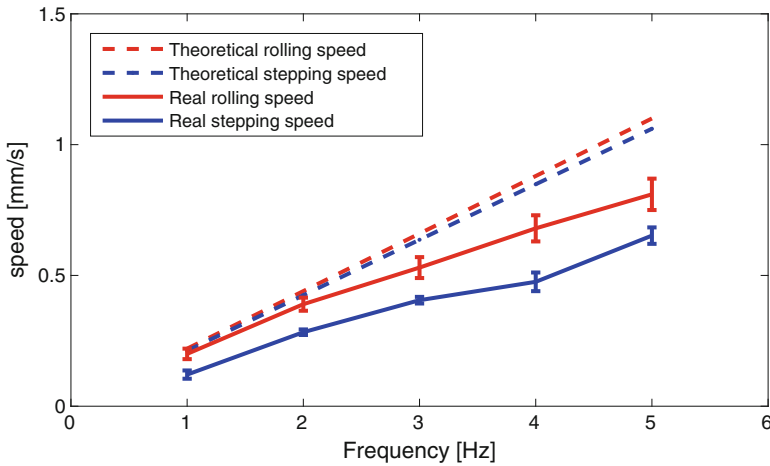
Fig. 4.6 Theoretical (*dashed lines*) and real (*solid lines*) forward speed for the rolling (*red lines*) and stepping (*blue lines*) locomotion technique. The real speed is obtained with water (rolling) and isopropanol (stepping) as liquid environment. For the stepping locomotion technique the microrobot's length is $l = 300\,\mu\text{m}$, the stepping angle $\phi = 45\,^\circ$ and the tilt angle of the magnetic field $\beta = 40\,^\circ$. Each measurement is averaged over three experiments and the error bar represents one standard deviation

β of the magnetic field, and was set to $\beta = 40\,^\circ$. The deviation of the experimental data from this ideal forward velocity indicates slip of the microrobot, as will be explained in the next section.

3.4 Evaluation

Figure 4.6 shows the theoretical (dashed lines) and real (solid lines) forward speed for the rolling (red lines) and stepping (blue lines) locomotion technique. For the rolling experiments a microrobot of type RL6N10 is used, while for the stepping

experiments a microrobot of type RW2L3N3 is used. The stepping locomotion is evaluated with a stepping angle of $\phi = 45°$ and a tilt angle of the magnetic field $\beta = 40°$. Because the experimental velocity is lower than the theoretical speed, slip of the microrobot limits the forward speed. One reason for this slip is the viscosity of the surrounding liquid, i.e. when the viscosity is higher this leads to more slip and a lower velocity [20]. Additionally, as forward motion is achieved due to the frictional contact between the microrobot and the surface, the roughness of the supporting surface plays a role as well. When the surface roughness is high, the microrobot is expected to have more traction than for a smooth surface, and the resulting forward speed will be higher. These explanations hold for both the rolling and stepping actuation technique. The motion of the microrobots was evaluated in water (rolling) and isopropanol (stepping) on a glass substrate.

4 Object Manipulation

The properties of the environment prescribe which object manipulation techniques are possible. Considering a liquid environment, the density and viscosity of the liquid, as well as the density of the object, dictate whether an object can be lifted or not. For example, the density of polystyrene ($1.05 \, \text{kg m}^{-3}$) is higher than the density of isopropanol ($0.79 \, \text{kg m}^{-3}$, at room temperature). This ensures a polystyrene bead will sink in isopropanol due to gravity, but can still be lifted up due to a generated fluid flow. This delicate balance between the properties of the object and the surrounding fluid demonstrates the importance of environmental conditions and enables two manipulation capabilities. First, manipulation via direct contact is shown, where the microrobot pushes an object. Second, the motion of the rolling microrobot can generate a local fluidic vortex which enables a noncontact manipulation technique.

4.1 Contact Manipulation

Object manipulation by pushing is a common strategy for microrobots and is possible when the object to be manipulated is heavy enough as not to be lifted by the fluid flow and light enough such that the microrobot can actually move it. This means that the apparent weight of the object, and the electrostatic and the Van der Waals forces due to the interaction of the object with the surface, must be lower than the force applied by pushing. Figure 4.7 shows a microrobot of type RW2L3N3 ($300 \, \mu\text{m}$ wide) pushing a micro-object ($200 \times 350 \, \mu\text{m}$) that is triangular shaped and is fabricated from SU-8 (density $1.2 \, \text{kg m}^{-3}$). Figure 4.8 shows a microrobot of type RL3N5 pushing four micro-objects (SU-8 triangles, $200 \times 350 \, \mu\text{m}$) such that they densely assemble into a narrow channel ($750 \, \mu\text{m}$ wide). Pushing manipulation is more effective with a rolling microrobot than with a stepping microrobot. This is because the stepping actuation technique achieves less stable motion than the
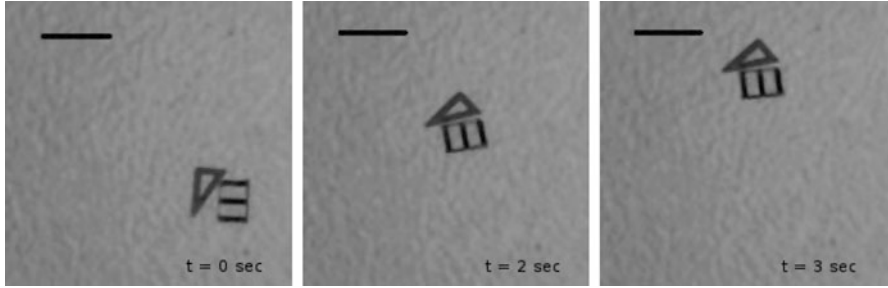
Fig. 4.7 Microscopic view of object manipulation via pushing. The microrobot of type RW2L3N3 (300 μm wide) pushes a micro-object (SU-8 triangle, 200×350 μm). The black scale bar is 500 μm long

Fig. 4.8 Microscopic view of object manipulation via pushing. The microrobot of type RL3N5 (300 μm wide) assembles four micro-objects (SU-8 triangles, 200 × 350 μm) densely packed into a narrow channel (750 μm wide) within 2 min. The black scale bar is 500 μm long

rolling technique, and the orientation of the device can be better controlled with rolling. The orientation of the stepping microrobot is continuously stepping in order to move forward which works against a continuous and stable pushing contact.

4.2 Noncontact Manipulation

The lifting effect of the fluid flow in front of the microrobot and near the surface occurs due to the no-slip condition of the fluid at the surface (shear flow) and causes objects to be lifted up from the surface. This fluidic lifting effect is local, as objects further away from the robot are not affected. The fluidic vortex above the microrobot is generated due to the robot's rotation and the fluid's no-slip condition at the surface. Both effects complement each other, and are possible due to the low Reynolds number environment [24]. Figure 4.9 shows a microrobot of type RL6N10 (600 μm wide) lifting and transporting a polystyrene bead that is 100 μm in diameter.
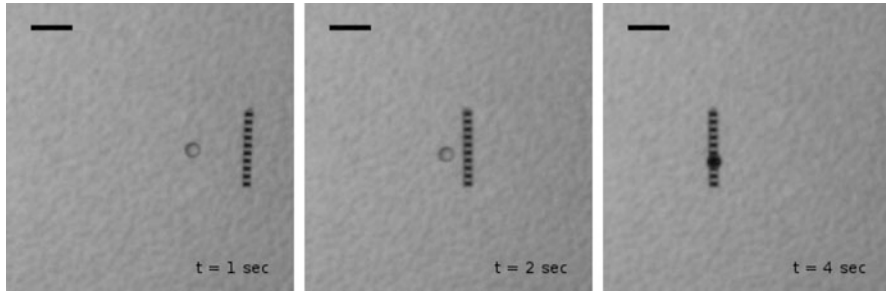
Fig. 4.9 Microscopic view of object manipulation via fluidic trapping. The microrobot of type RL6N10 (600 μm wide) approaches (*left*), lifts (*middle*) and transports (*right*) a polystyrene bead with 100 μm in diameter. The black scale bar is 300 μm long

5 Conclusions

This chapter presented the design and capabilities of a family of microrobots. While the fabrication process is identical for all designs, the method for actuation and the object manipulation capabilities are specific for each particular design. Each microrobot design has embedded soft-magnetic posts that align to a magnetic field. Depending on the arrangement of these posts the microrobot can be actuated by either a rotating magnetic field or a stepping magnetic field. Consequently, each actuation technique enables a different object manipulation techniques (i.e. contact or noncontact manipulation). In particular, a rotating magnetic field enables a local vortex above the robot that traps objects for transportation. Moreover, a pushing manipulation technique can be achieved for both the rotating and stepping magnetic field actuation. Experimental results demonstrate all mentioned techniques for different microrobotic designs.

References

1. Abbott J, Nagy Z, Beyeler F, Nelson B (2007) Robotics in the small, part I: microbotics. IEEE Robot Autom Mag 14(2):92–103
2. Purcell E (1977) Life at low Reynolds number. Am J Phys 45(1):3–11
3. Abbott JJ, Peyer KE, Cosentino Lagomarsino M, Zhang L, Dong LX, Nelson BJ (2009) How should microrobots swim? Int J Robot Res 28(11–12):1434–1447
4. Sun Y, Liu X (2015) Micro- and nanomanipulation tools. Wiley, New York
5. Bouchebout S, Bolopion A, Abrahamians J-O, Régnier S (2012) An overview of multiple DoF magnetic actuated micro-robots. J Micro-Nano Mechatronics 7(4):97–113
6. Xu T, Yu J, Yan X, Choi H, Zhang L (2015) Magnetic actuation based motion control for microrobots: an overview. Micromachines 6:1346–1364
7. Banerjee A, Gupta S (2013) Research in automated planning and control for micromanipulation. IEEE Trans Autom Sci Eng 10(3):485–495

8. Savia M, Koivo H (2009) Contact micromanipulation - survey of strategies. IEEE/ASME Trans Mechatronics 14(4):504–514
9. Kummer M, Abbott J, Kratochvil B, Borer R, Sengul A, Nelson B (2010) Octomag: An electromagnetic system for 5-DOF wireless micromanipulation. IEEE Trans Robot 26(6), 1006–1017
10. Bergeles C, Kratochvil B, Nelson B (2012) Visually servoing magnetic intraocular microdevices. IEEE Trans Robot 28(4):798–809
11. Frutiger DR, Vollmers K, Kratochvil BE, Nelson BJ (2009) Small, fast, and under control: wireless resonant magnetic micro-agents. Int J Robot Res 29:613–636.
12. Pawashe C, Floyd S, Diller E, Sitti M (2012) Two-dimensional autonomous microparticle manipulation strategies for magnetic microrobots in fluidic environments. IEEE Trans Robot 28(2):467–477
13. Steager EB, Sakar MS, Magee C, Kennedy M, Cowley A, Kumar V (2013) Automated biomanipulation of single cells using magnetic microrobots. Int J Robot Res 32:346–359
14. Nelson BJ, Kaliakatsos IK, Abbott JJ (2010) Microrobots for minimally invasive medicine. Ann Rev Biomed Eng 12:55–85
15. Crane NB, Onen O, Carballo J, Ni Q, Guldiken R (2012) Fluidic assembly at the microscale: progress and prospects. Microfluid Nanofluid 14(3):383–419
16. Floyd S, Pawashe C, Sitti M (2009) Two-dimensional contact and noncontact micromanipulation in liquid using an untethered mobile magnetic microrobot. IEEE Trans Robot 25(6):1332–1342
17. Pieters RS, Tung H-W, Charreyron S, Sargent SF, Nelson BJ (2015) RodBot: a rolling microrobot for micromanipulation. In: Proceedings of IEEE international conference on robotics and automation (ICRA), pp 4042–4047
18. Happel J, Brenner H (1983) Low Reynolds number hydrodynamics: with special applications to particulate media. Martinus Nijhoff, The Hague
19. Bhushan B (1998) Handbook of micro/nano tribology. CRC Press, Boca Raton
20. Tung H-W, Peyer KE, Sargent DF, Nelson BJ (2013) Noncontact manipulation using a transversely magnetized rolling robot. Appl Phys Lett 103(11):114101
21. Tung H-W, Sargent DF, Nelson BJ (2014) Protein crystal harvesting using the RodBot - a wireless, mobile microrobot. J Appl Crystallogr 4:692–700
22. Ergeneman O, Sivaraman KM, Pané S, Pellicer E, Teleki A, Hirt AM, Baró MD, Nelson BJ (2011) Morphology, structure and magnetic properties of cobalt nickel films obtained from acidic electrolytes containing glycine. Electrochim Acta 56(3):1399–1408
23. Kratochvil BE, Kummer MP, Erni S, Borer R, Frutiger DR, Schürle, S, Nelson BJ (2014) MiniMag: a hemispherical electromagnetic system for 5-DOF wireless micromanipulation. In Khatib O, Kumar V, Sukhatme G (eds) Experimental Robotics. Springer Tracts in Advanced Robotics, vol 79. Springer, Berlin, pp 317–329
24. Merlen A, Frankiewicz C (2011) Cylinder rolling on a wall at low Reynolds numbers. J Fluid Mech 685:461–494

Chapter 5
Integrating Smart Mobile Devices for Immersive Interaction and Control of Physical Systems: A Cyber-Physical Approach

Jared A. Frank and Vikram Kapila

Abstract The embedded technologies integrated into smart mobile devices are becoming increasingly more powerful and being applied to solve disparate societal problems in unprecedented new ways. Billions of smartphones and tablet computers have already reshaped the daily lives of users, and efforts are currently underway to introduce mobile devices to some of the most remote and impoverished areas of the world. With an ever-expanding list of sensors and features, smartphones and tablet computers are now more capable than ever of enhancing not only our interactions with software and with each other, but with the physical world as well. To utilize smart mobile devices at the center of rich human-in-the-loop cyber-physical systems, their sensing, storage, computation, and communication (SSCC) capabilities must be examined from a mechatronics perspective rather than the contexts in which they are conventionally treated (e.g., messaging, surfing the web, playing games, navigation, and social networking). In this chapter, we discuss how state-of-the-art mobile technologies may be integrated into human-in-the-loop cyber-physical systems and exploited to provide natural mappings for remote interactions with such systems. A demonstrative example is used to show how an intuitive metaphor is uncovered for performing a balancing task through the teleoperation of a ball and beam test-bed.

1 Introduction

Our modern technological age abounds with user interfaces to interact with a myriad of technologies that we encounter in our daily lives (e.g., kitchen appliances, consumer electronics, office machines, automobiles, etc.). When a user interface

J.A. Frank • V. Kapila (✉)
NYU Tandon School of Engineering, Brooklyn, NY 11201, USA
e-mail: jared.alan@nyu.edu; vkapila@nyu.edu

is intuitive, users are able to draw from their prior experiences and familiarity to immediately use the interface effectively. This allows the interface to become transparent to the user and the user's attention to be directed towards performing the task rather than trying to learn how the interface is used. That is, when interfaces are made intuitive, users can perform tasks with ease, comfort, and delight.

Besides the traditional button- and knob-based interfaces that we encounter on a daily basis, people now interact with graphical interfaces that exist on mobile devices. Not only are these interfaces used everyday by the public to accomplish the tasks of the internet age (e.g., accessing email, searching the internet, social networking, listening to music, and checking the weather), the inherent mobility of smartphones and tablets has led to a growing list of novel interactive activities designed to socially connect people (e.g., tweeting, snapchatting, location-based dating, etc.). However, the overwhelming number of these interactions occurs in cyberspace, without any physical element. That is, such interfaces typically enable people to virtually interact either with each other or with digital information rather than with physical systems like machines.

Society is experiencing an escalation in the complexity of engineering systems. Machines that were once simple are growing ever more complex and occupying new application domains. In particular, a host of robotic technologies that has already penetrated industry is expected to appear both in the home and at the workplace in the near future. Although sophisticated machines like robots have been remotely controlled, or teleoperated, by trained technicians for many years, their painless adoption by the general public will require nontechnical and inexperienced consumers to be provided with the most intuitive user interfaces possible. With the steady advancement of embedded sensing, storage, computation, and communication (SSCC) capabilities, smart mobile devices like smartphones and tablets are more capable than ever to serve as platforms for providing immersive interactions with such systems. Moreover, because of the already significant popularity and familiarity of smart mobile devices, such interfaces may be accessible to users with little to no additional cost or training.

Human-in-the-loop cyber-physical systems, in which user input, computation, communication, and control of physical dynamics are more intimately interconnected than in traditional teleoperation systems, have the potential to augment the user's interaction with the physical world [32]. However, the incorporation of a mobile device as a component in such a system is not a trivial undertaking. It involves the synergistic integration of the mobile device hardware and software towards a high-level objective that goes beyond the conventional purposes for which the embedded technologies have been designed and used. New challenges include how the SSCC capacity of mobile devices can be used to (1) capture and map user behavior to desired behavior of the physical system to support the creation of intuitive metaphors, and (2) capture and display the state of the physical system to the user to support situational awareness. These considerations are essential in achieving a level of immersion that enables the user to effectively transfer acquired skills in interacting with mobile interfaces to successfully operate the physical system. Note that to accomplish such rich interactions between human users and
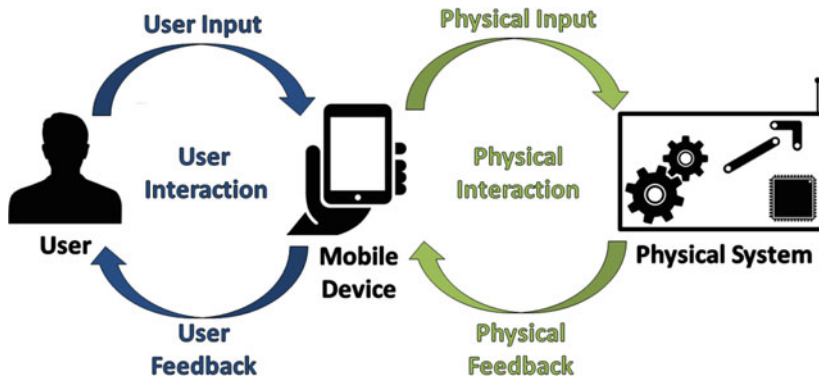
Fig. 5.1 Architecture of a cyber-physical system that enhances user interaction with a physical system using a mobile device

physical systems, the mobile device occupies the center of two interdependent cyber-physical loops (see Fig. 5.1). In this architecture, the properties of the physical system (e.g., stability and performance of the physical process) and conditions of the user experience (e.g., responsiveness and usability of the user interface) can become intimately coupled and dependent on the speed and reliability with which the mobile device processes, stores, and communicates information between the user and the physical system. Therefore, both the physical and user interaction loops impose demands on the hardware and software of the mobile device as it is used to close the two loops.

After the physical process is constructed, the mobile application is developed, and the functionality of the components is tested, a successful interaction is contingent upon the usability of the system. With a human-in-the-loop system, the outcome is difficult, if not impossible, to predict *exactly* by means of theory, computation, or simulation. Therefore, the performance of the closed-loop system must be expressed in terms of quantitative metrics and studies must be conducted in which users interact with the system. In this chapter, the development of a ball and beam system for studying rich immersive interactions using a smartphone is described. The work presented in this chapter is based on [14].

The chapter is organized as follows. In Sect. 2, the motivation behind the integration of smart mobile devices to build human-in-the-loop cyber-physical systems is discussed. Section 3 describes the hardware technologies of mobile devices that make them suitable for interaction with and control of physical systems. Then, Sect. 4 discusses the software aspects, including mobile operating systems and mobile application development, that allow for the immersive interaction with and control of physical systems using smart mobile devices. Next, Sect. 5 discusses the development of a human-in-the-loop cyber-physical system in which a mobile application is used to uncover an intuitive interaction metaphor for performing a remote balancing task through the direct control of a ball and beam test-bed. Finally, Sect. 6 offers concluding remarks and future directions.

2   Motivation

The release of the first smartphones with diverse sets of embedded sensors in 2007 was soon accompanied with the development of mobile interfaces that utilize those sensors to monitor users' physical activities, heart rate, and driving style [5]. These mobile applications have brought attention to the use of mobile devices as handheld measurement systems and have come to monitor much more than just the user of the device, impacting areas such as healthcare, transportation, and environmental monitoring [22, 25].

Almost immediately after exploring the measurement applications of smart mobile devices, development began of an even more compelling class of mobile applications that allow users to operate physical systems. Of these systems, the overwhelming majority are classified as robots or unmanned vehicles and used to extend the user's physical influence in a remote environment [1, 3, 8, 16, 23, 34]. These efforts are in many ways a continuation of telerobotics work with earlier generations of mobile technologies, such as personal digital assistants [11, 38]. Regardless of the interface hardware or software, users traditionally interacted with physical systems through a direct control approach, in which the physical system exhibits little to no autonomy and thus all the monitoring, decision making, and planning are performed by the user. Although this scheme was quick and easy to implement, direct control presented several problems in time-critical applications with significantly delayed communications that were resolved with the introduction of supervisory control schemes [10, 33]. Supervisory control, in which the physical system remains semi-autonomous and thus relatively stable in the face of communication delays, permits the user to intermittently monitor and intervene with the control of the physical system, shifting the focus to remotely commanding systems at a high level of abstraction.

Advances in the SSCC capabilities of mobile devices like smartphones and tablets have led to interest in their use as platforms for cyber-physical systems [2, 37]. Unlike traditional embedded systems, these systems give rise to a number of conceivable architectures in which the roles of sensing and computation are distributed amongst a network of heterogeneous interacting elements [21]. Mobile interfaces designed to operate as part of human-in-the-loop cyber-physical systems may enable immersive interactive experiences with physical systems using any desired control scheme. Regardless of the chosen control scheme, low-level interrelationships will exist between the SSCC capabilities of the mobile device, characteristics of the physical dynamics of the engineering system, and aspects of the user experience. These interrelationships are of particular interest in direct control architectures, since although the direct control scheme is well suited for applications that require real-time human decision making and that can support low-delay communications, their effects can be destabilizing. By addressing these challenges on mobile platforms, the performance of more challenging teleoperation tasks with relatively unstable systems may be possible, such as the balancing task demonstrated in this chapter.

Moving some of the computational and communicative load onto the mobile device can potentially reduce demands such as the size, cost, complexity, and energy usage of the engineering system being controlled. However, this comes at the expense of increasing the sensitivity of the system to user actions and the network performance, and adding an additional layer of complexity to the user interface design problem. In particular, mobile interfaces must provide users with services such as intuitive interaction metaphors, a variety of feedback modalities, and a bidirectional non-blocking communication routine for exchanging information with the physical system, while remaining responsive and meeting timing constraints to maintain the stability of the physical process. To achieve this goal, the hardware and software components of the devices must be assessed with respect to the roles they are expected to serve in these systems.

3   Smart Device Hardware

Although the performance and user experience associated with utilizing smart mobile devices in the interaction and control of physical systems depend on many factors, the SSCC capabilities of the mobile device are essential to the successful operation of the system. The requirements of the hardware will vary depending on the physical process to be controlled and in some cases may be beyond the capabilities of the current generation of devices. For this reason, developers should stay up-to-date on the capabilities and limitations of the device hardware, as they continue to advance at an accelerated rate.

3.1   Sensing

To utilize a smart mobile device to interact and control physical systems, the ability to accurately capture information from the user or from the physical system is of paramount importance. As is the trend in electronics, sensors continue to be made smaller, more powerful, and affordable. This is good news, since sensors have become an integral part of shaping the mobile experience. Many phones now contain sensors to measure everything from the location, motion, and orientation of the device; ambient temperature, light, pressure, and humidity; as well as the proximity, fingerprint, footsteps, voice, and heart rate of the user. Although not all sensors are embedded in all devices, most if not all of the relevant technologies for extracting users' gestures are now standard on modern devices.

Touchscreen   When it comes to user interaction, the device screen plays a pivotal role. Not only must it be used to provide the main source of feedback to the user to support situational awareness during the interaction, but the screen also serves as one of the most important sources of user input. Unlike most touchscreens, the touchscreens of modern smartphones and tablets are based on capacitive sensing technology. This means that touchscreens have not only become more sensitive,

responsive, and accurate than previous generations of resistive screens, but their displays have become much sharper as well. This has resulted in the creation of rich touch and multi-touch interactions with mobile interfaces.

Inertial Sensors People have been using forms of gestures to communicate non-verbal messages to each other since the beginning of history. By using inertial sensors such as 3-axis accelerometers, gyroscopes, and magnetometers, mobile device manufacturers have made it possible to accurately capture the movements of devices (e.g., tilts, shakes, rotations, and swings) and to recognize gestures from movement data. Moreover, the ability to recognize gestures from movement data empowers mobile devices to act as extensions of the user and is beginning to play an important role in creating a range of new promising interaction possibilities. With the integration of dedicated motion coprocessors, devices can now collect, process, and store motion data without burdening the application processor.

3.2 Storage

The need to collect and process data from the user and from engineering systems on smart mobile devices necessitates the ability to store and work with large amounts of data. With mobile devices that contain up to 128 GB in flash memory and as much as 4 GB of RAM, this is not a problem on current platforms.

3.3 Computation

Many aspects of the performance and user experience associated with interacting and controlling a physical system from a smart mobile device are closely linked to the computational power of the device. During the interaction, the device is expected to collect and store data from a multitude of sensors, to execute a multitude of operations on the data, to render 2D or even 3D graphics to display relevant data to the user, and to communicate information with the physical process, oftentimes simultaneously and many times per second. In this way, the processor plays a central role in satisfying both functionality and usability requirements of the system. Manufacturers are now producing systems-on-a-chip (SoCs), in which dual- or even quad-core processors, running at rates several times faster than their first released models, are integrated with graphics processors, coprocessors, GPS, cellular modems, and memory on a single board. The SoC not only allows the size of the device to be reduced, but also dramatically improves speed and power usage. In fact, the computational power of mobile devices has quickly caught up to the level of desktop PCs and has been responsible for a shift towards a mobile computing era in which smartphones and tablets are the central computing devices in people's lives.

3.4 Communication

If the interactions discussed in this chapter were just between the user and the mobile interface, then simply considering the sensing, storage, and computation of the mobile device would have been sufficient. However, wireless connectivity is necessary to communicate data between the mobile device and the engineering system. Support for the latest generations of mobile, Wi-Fi 802.11, bluetooth, infrared, and near-field communications enables mobile devices to access and share information with other devices at incredible speeds and over a variety of ranges. This communication will introduce small delays to the closed loop, which must be minimized as they negatively affect the stability and performance of the physical system as well as the responsiveness of the interface.

4 Software

The purpose of mobile device software is to harness the embedded hardware in providing rich intuitive user experiences. This includes the mobile operating system, which provides the hardware drivers, the libraries, frameworks, and application programming interfaces (API) so that developed applications have access to the sensors, features, and data available on the device. It also includes the mobile applications themselves, which make use of the available sensors, features, and data to provide specialized functionality to the user. However, the design of interfaces that enable users to interact with physical systems presents an additional layer of complexity: the user must directly interact with the application on the mobile device to indirectly interact with the physical system. In other words, mobile interfaces must provide natural mappings from the user's actions on the interface to the commands that will be communicated to influence the state of the physical system. In addition, the interface must provide clear feedback that can be used to interpret the state of the physical system. It is natural to expect that the choice of mappings and feedback will be dictated by the nature of the physical system. For assistance, developers must leverage the research and heuristics that have been gathered on the design of teleoperation systems [33] as well as on the design of mobile interfaces in general [31, 36].

A well-designed interface must reinforce users' assumptions and expectations of how the interface is to be used and their mental models of how the physical system will behave in response to their actions. To avoid overwhelming users, the low-level details of the system ought to be hidden from the user and replaced with high-level metaphors that may be familiar to the user [30]. The choice of interaction metaphors may be inspired from the physical nature of the system to give the user the most natural experience when using a mobile interface (e.g., exploiting a spatial analogy wherein a physical object moves up or down when the user moves an interactive object on the interface up or down, respectively). These interaction metaphors can be

realized with gestures extracted from the touchscreen or from inertial sensors. Such metaphors have been shown to be effective in designing interactions using the same sensors and features on tangible user interfaces for human–robot interaction [18], in gaming technology to build natural interfaces that enrich the user interaction with desktop computers [12], and game controllers for introducing the elderly to video games that keep them healthy through physical activity [17]. These applications demonstrate that user interfaces with intuitive metaphors have the potential to make remote interaction with physical systems more natural.

The design of metaphors for interacting with a particular system poses an open-ended, creative design challenge that lacks a unique solution. Thus, it is important to rigorously examine each of the metaphors that may provide a natural mapping. This requires a comparative study to determine the interaction metaphor that not only feels most natural to users but that also yields acceptable performance of the physical process.

5 Putting It All Together: Control and Interaction with a Ball and Beam Test-Bed

We will now discuss the development of a human-in-the-loop cyber-physical system that implements a smartphone to enable immersive interaction with a ball and beam test-bed. During the interaction, the control of the angular orientation of the motorized beam is maintained by a laboratory station built around a PC-based data acquisition and control board (DACB). The goal is to use the mobile interface on the smartphone to monitor and command the orientation of the beam such that the ball remains balanced at the center of the beam. In other words, while the stability of the beam dynamics is ensured by the laboratory station, the ball dynamics are intimately related with user behavior, computation, and communication in the system. To determine the most effective metaphor for completing this task, a user study is performed wherein participants are asked to perform the task using several designed metaphors. A survey is conducted with the participants to determine the user satisfaction with each metaphor. The experimental data shows that the preferred metaphor of the participants is the one in which the smartphone is tilted to mimic the desired tilting of the beam, and is the same one which yields significantly better task performance.

5.1 System Description

The ball and beam test-bed consists of a DC-motor, a 0.5-m long lexan plastic beam, and a smooth 1 in. (0.0254 m) diameter steel ball. The output shaft of the DC-motor is attached to a gearbox to produce sufficient torque to drive the beam mounted to the output shaft. Attached to the shafts of the gearbox are a potentiometer and an absolute encoder for measuring angular displacement and a tachometer for measuring
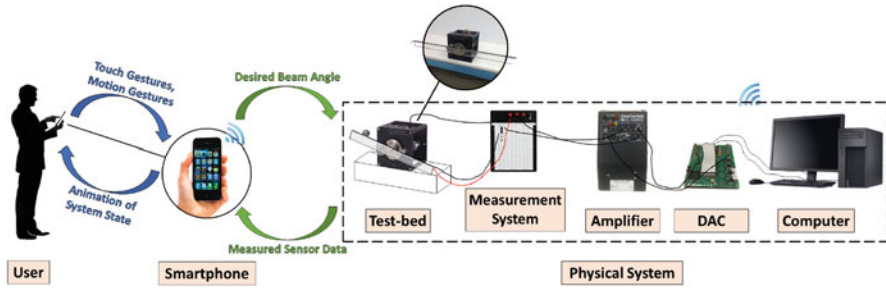
Fig. 5.2 The human-in-the-loop cyber-physical system for interacting with the ball and beam test-bed using a smartphone

angular velocity. A setup with conductive and resistive strips attached to a track on the beam, and associated electronics, is used to measure the position of the ball along the beam. A computer running MATLAB/Simulink wirelessly communicates with the mobile device and stabilizes the DC-motor to the desired orientation using an optimal control algorithm (see Appendix), a PC-based DACB, and a power amplifier. Figure 5.2 shows the complete human-in-the-loop cyber-physical system, its components, and the data communicated between its major constituents.

5.1.1   Measurement System

To control the angle of the beam attached to the output shaft, measurements of the angular position and velocity of the output shaft are, respectively, acquired using the encoder and a tachometer mounted to the test-bed. For the user to be provided with an animated display of the state of the system on the mobile interface for situational awareness, and to evaluate the performance of the balancing task, a measurement system is designed and integrated with the test-bed to supply measurements of the position of the ball to the desktop computer. The circuit shown in Fig. 5.3 is connected to the desktop computer via the DAC board. More information about the setup of the DAC board is found in [20]. A carbon-based resistive strip and a copper-based conductive strip are used to create a linear potentiometer on the beam with the ball acting as the wiper. Since the analog channels of the DAC board have a range of $\pm 10V$, an operational amplifier is used to bring the output of the linear potentiometer to this range. Two 10 kW potentiometers are used in the operational amplifier circuit to trim the gain and offset of the output voltage.

5.2   *User Interface Design*

The smartphone-based user interface for interacting with the ball and beam test-bed is developed to run on an Apple iPhone 5, with iOS as the mobile operating system. When the application is started, the user first encounters the main menu
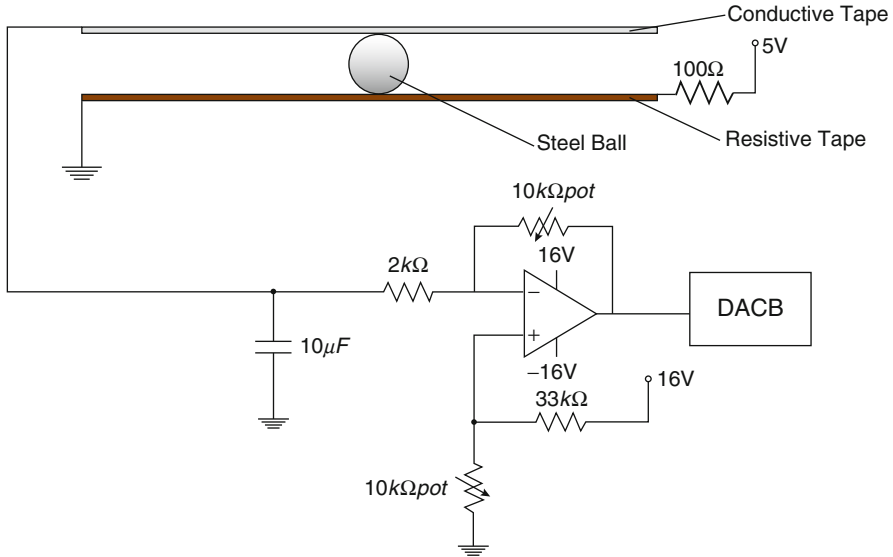
Fig. 5.3 The system for measuring the position of the ball along the length of the beam
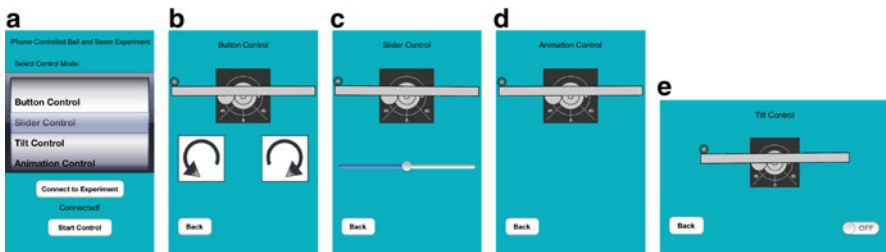
Fig. 5.4 Screenshots of the smartphone application, including the (**a**) main menu and the interface modes for controlling the ball and beam test-bed using (**b**) buttons, (**c**) slider, (**d**) direct manipulation of the animation, and (**e**) tilt

screen (see Fig. 5.4a), where she may select an interface mode to interact with the test-bed and press a button to establish a wireless connection with the test-bed using Wi-Fi 802.11. To allow for ease of navigation, the layout of the interface is inspired by standard utility applications that users are familiar with, while the design of the interactive interface is modeled after the look and feel of game applications, the most widely downloaded class of mobile applications [29]. A rich two-dimensional graphical animation of the ball and beam test-bed is included in the interface, and updates in real time as the application receives sensor data from the test-bed. Inputs from the touchscreen and from the inertial sensors on the device are used in designing the interface modes. Specifically, four different interface modes are made available to interact with the ball and beam test-bed, including tapping virtual buttons, dragging a virtual slider, directly manipulating the animation by tapping and dragging on the virtual beam, and tilting the device as if it were the beam.

These interface modes resulted from (1) a narrow selection process in which more than twice the number of interactive alternatives was evaluated, followed by (2) a series of refinements in which the final values were chosen for parameters governing the mappings. These steps helped to discover usability problems associated with the interface, make predictions concerning tendencies in user behavior, and establish a set of context-specific heuristics to provide a consistent level of usability for future applications [24]. To conduct a study in which the design choices can be validated and the user interface modes can be compared, each mode was used independently and provided in separate views (Fig. 5.4b–e).

Buttons With the first user-selectable interface mode (Fig. 5.4b), two buttons are pressed to drive the beam clockwise and counterclockwise. Several alternative mappings from the button presses to test-bed commands were considered in the design of this mode. Inspiration for the mappings was drawn from the developers' experience with teleoperation systems and video games. Some of the mappings were based on distinct taps on the buttons and others required holding the buttons down. Finite angular displacement commands, velocity commands, and acceleration commands to the beam in the directions indicated by the symbols on the buttons were considered. Ultimately two of the mappings were combined to enable the user to tap the buttons to rotate the beam in increments of $1°$ at a time or hold the buttons to drive the beam at a quick but constant angular velocity ($80°$/sec). This gave the user the ability to make fine adjustments to the beam angle as well as liberally drive the beam depending on the situation.

Slider In the second user-selectable interface mode (Fig. 5.4c), a slider is dragged across the screen to command the orientation of the beam. In this case, the major design decisions focus on the orientation of the slider (horizontal versus vertical), and the sensitivity of the mapping, which is dictated by the length of the slider and the range of values covered by the slider. Preliminary evaluations revealed that to balance the ball, the beam needs to remain in a region $\pm 10°$ from its horizontal orientation. Also, users tend to make larger sweeps of the slider as opposed to smaller ones. Thus, a slider that spans the entire width of the screen was used to command the beam to within the appropriate angular region ($\pm 10°$), providing users the needed sensitivity to make quick, fine adjustments to beam orientation.

Animation In the third user-selectable interface mode (Fig. 5.4d), the user can interact with the beam by directly interacting with a one-ninth scale animation of the beam using simple gestures on the touchscreen. Several alternative mappings were considered when designing this direct manipulation mode, including tapping on the screen in order for the animated beam to align itself with the user's finger and dragging the animated beam to the desired orientation for reorienting the beam. The main issues encountered during the design of this interface mode included whether users would choose to tap or drag the animation and where on the screen they would perform these actions. Due to the resolutions of the screen and the beam animation, it is more difficult to command the orientation of the beam when interacting with the animation near the center of the beam. Thus, the distance in pixels between the

user's finger and the center of the beam is used to reduce the sensitivity as the user interacted closer to the center of the beam. Interestingly this solution of the usability problem produces a virtual moment arm effect. In fact, the results of preliminary evaluations reveal that many users' mental models already include this moment arm effect, as most users tend towards interacting with the animation at the ends of the beam. This interaction provides the user with the sensation that she is grabbing the actual beam by one end and tipping it upward or downward. Moreover, preliminary evaluations predicted that, in general, users tend to make large sweeping movements across the screen when required to manipulate the beam quickly to complete the exercise. Therefore the sensitivity of the interaction was lowered by reducing the ratio between the output (angle of the beam with respect to the horizontal) and the input (vertical finger motion in pixels) to $0.06\,°$/pixel. This allowed users to make fine adjustments with large drags of the fingers on the screen. Although it meant that users were not tapping exactly at the angle on the animation where they would like the actual beam to be oriented, this did not have a significant impact on users' ability to control the beam, since they mostly relied on visual feedback from the test-bed to orient the beam.

Tilt The fourth user-selectable interface mode (Fig. 5.4e) uses the device's built-in accelerometer sensor to command the orientation of the beam. This is based on the underlying assumption that the only acceleration experienced by the device accelerometer is due to gravity [9]. By tilting the mobile device, the user is provided with an interaction metaphor that gives her the sensation that she is actually tilting the beam itself. The tilt-based mode includes a toggle switch to turn the control of the beam on or off, to prevent commands from being sent to the test-bed prematurely until the user has correctly oriented the smartphone. The main design parameter of this interaction metaphor is the ratio between output (the angle of the beam with respect to the horizontal) and input (the angle of the mobile device). This ratio significantly impacts the sensitivity of the interaction. Originally, the ratio was designed to be 1:1; however, the results of preliminary evaluations indicated that users tend to exert large rotations on the device when required to manipulate the beam quickly to complete the exercise. Therefore, the sensitivity of the interaction is reduced to a ratio of 1:3 to improve performance. This allowed users to still make fine adjustments to beam orientation with generous motions of the device.

The mappings of the different interface modes for human–machine interactions presented above are not unique. It is certainly possible to explore alternative mappings and metaphors to control the system. For example, prior research has investigated additional modes, such as knobs and joysticks, for the control of mobile robots [26]. In addition, prior research has also investigated the use of vision as a modality to foster intuitive user interactions with educational laboratory test-beds [15]. However, the focus of this work is to illustrate how the hardware and software elements embedded in mobile platforms can be used in human-in-the-loop cyber-physical interactions and to lay a foundation for further research into designing such interactions. A longitudinal study, in which each of the modalities is iteratively improved upon, will be considered in future work.

5.3   User Study

To investigate the usability of the designed smartphone interface, validate the potential of utilizing smartphone hardware and software in systems such as the one described in this chapter, and determine the most effective interaction metaphor for completing the balancing task, a study is conducted with 21 participants (20 male, 1 female), all of similar age and technological experience (engineering students between 19 and 26 years old). After completing a preliminary survey, each participant uses each of the four interface modes three times (for a total of 12 trials per participant and 252 total trials) to control the orientation of the beam to balance the ball as close to the center of the beam as possible for 20 s. During the interaction, participants stand directly in front of the test-bed and have the choice to either look at the animation of the test-bed on the smartphone screen or to look at the actual test-bed. The order in which the participants use the interface modes is randomized to prevent any order effect on the results of the study. Specifically, the four interface modes yield 24 permutations for the order in which they can be used. Thus, no two of the 21 participants are assigned the same order. At the beginning of each trial, the ball begins at the left end of the beam. When the user is ready, she may use the interface to bring the ball towards the center of the beam, at which point the twenty-second timer is started. The task is completed once either the twenty seconds have ended or the ball has fallen from either end of the beam. As participants complete the exercise, the position of the ball is measured and recorded by the desktop computer in MATLAB. The outcome of the task (whether the ball remains balanced or not) is also recorded. The performance metrics are the score (1 if the ball remains balanced after twenty seconds and 0 otherwise) and the root mean squared deviation (RMSD) of the ball from the center of the beam for those trials that are successful. Of the users who successfully keep the ball balanced, those who have a more difficult time balancing the ball yield larger values for the RMSD and those who have an easier time keeping the ball balanced near the center yield smaller values for the RMSD. After the participants complete the task with each of the interface modes, they complete a final survey in which they indicate their level of disagreement/agreement (on a scale of 1–5) with statements concerning the ease of use and satisfaction with the interface. Participants are encouraged to provide written comments and suggestions, which are used to determine which interface mode each user preferred. The experiment is conducted with each participant according to the following protocol:

1. The participant completes a preliminary survey.
2. The participant is introduced to the smartphone application.
3. The order in which the participant will use the interface modes is generated.
4. The participant has three twenty-second attempts to balance the ball on the beam using the first interface mode.
5. Step 4 is repeated for the remaining three modes.
6. The participant is given the final survey which asks about participant's experience in using each of the modes.

As discussed in [27], user input to a machine can be in a continuous or a discrete format. Since the slider, animation, and tilt interaction modes yield continuous format data and the button mode yields discrete format data, sensitivities of different modes producing data of different formats are not amenable to comparison. Moreover, due to differences in nature between modalities (i.e., tilting the device or dragging a finger on the screen), sensitivities of different modes producing data of same formats are also not amenable to comparison. Thus, in this work, the sensitivity of each mode has been treated as an adjustable parameter that is critical to the performance and acceptance of a particular interaction metaphor, but that needs to be studied independently through user testing.

5.3.1 Quantitative Results

Each of the 21 participants in this study is successful in balancing the ball on the beam at least one time. However, eight of the participants balance the ball successfully by using only one of the interface modes, and seven of these eight participants are successful with only the tilt mode. Figure 5.5a shows the percentage of participants who are successful with each of the interface modes. It is important to note that none of the 21 participants successfully use the button mode even once to complete the exercise. Regarding the remaining three modes, percentages of successful participants are as follows: 28.6 % for the slider mode, 19 % for the animation mode, and 71.4 % for the tilt mode. Therefore, the tilt mode is undoubtedly superior to the other three mode in terms of the success rate of participants. Of the 252 trials, 75 are successful. Figure 5.5b shows the percentage of these 75 successful trials that each of the four interface modes are responsible for. Note that 60 % of the successful trials result from the use of the tilt mode, which is higher than the combined number of successful trials resulting from the animation and slider interface modes.
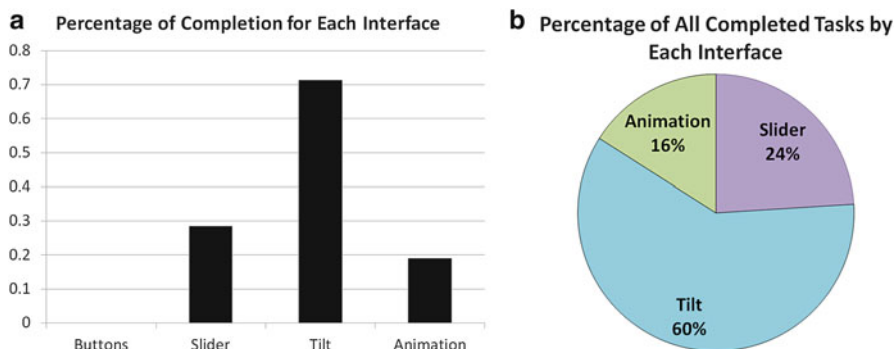

Fig. 5.5 Success results, including (**a**) percentage of successful participants for each mode and (**b**) percentage each mode is responsible for the total number of completed trials
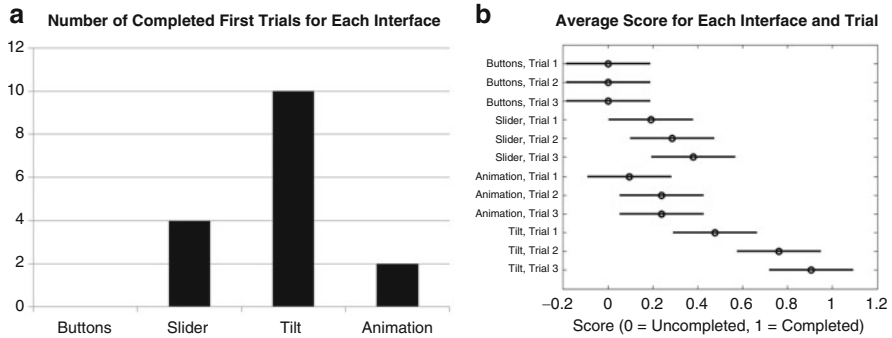
Fig. 5.6 Intuitiveness and learnability results, including (**a**) number of successful first trials for each mode and (**b**) multiple comparison test results for the effects of trial number and interface mode on score

To investigate the intuitiveness of each of the interaction metaphors, the number of times that participants successfully complete the balancing task in the first trial is recorded for each interface mode. Figure 5.6a shows that the tilt mode allows approximately half the participants to successfully balance the ball on the beam with no prior experience using the interface mode. An analysis of variance (ANOVA) test on the collected data reveals that the effect of the interface mode used on the average scores of the participants has a p-value of $p \ll 10^{-5}$, much less than the critical value of 0.05 for a 95 % confidence level. However, the effect of the trial number (first trial, second trial, or third trial) also has a significant effect on the average scores of the participants, with a p-value of 3.3×10^{-3}. No interaction between these two effects is found. To further investigate how the interface modes and trial numbers yield significant differences in score, a multiple comparison test is performed. The results of the multiple comparison test are shown in Fig. 5.6b. Note that the bars displayed in the figure are not error bars calculated from the variance in the data, but rather equal width intervals calculated using a modified Tukey–Kramer procedure [19] and used by MATLAB for purposes of statistical comparisons. If any two intervals are disjoint, then there is a significant difference between the two associated data points. Although participants perform better on average after successive trials using each interface mode (except for the button mode), the tilt mode is the only one of the four interface modes for which participants exhibit a significant increase in skill between their first and last trials. Participants are more likely to complete the task successfully on average with the tilt mode on their first attempt than with any trial using any other interface mode. Finally, the number of participants who complete the balancing task on their second trial with the tilt interface is statistically greater than the number of successful participants in any trial with any other interface.

To further investigate the performance of the participants, the RMSD was measured for each successfully completed trial. The results of an ANOVA test indicate that the interface mode used has a significant effect on the average RMSD
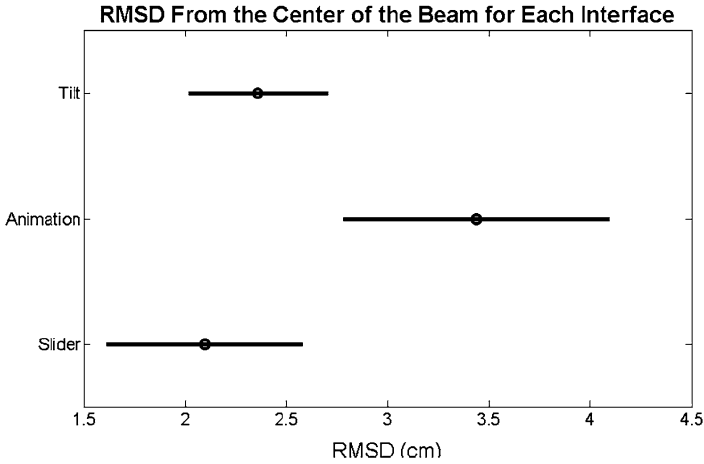
Fig. 5.7 Root mean squared deviation (RMSD) from the center of the beam for each interface

value associated with a successful trial ($p = 1.73 \times 10^{-2}$). A multiple comparison test performed to investigate this effect further yields the results shown in Fig. 5.7. These results indicate that participants using the tilt mode and slider mode are able to balance the ball significantly closer to the center of the beam than the participants using the animation mode.

5.3.2 Qualitative Results

A survey with statements motivated by [35] is given to participants after they complete the interaction exercise. The survey asks participants to agree or disagree with the statements for each of the interface modes according to a 5-point Likert scale, with 1 indicating strong disagreement and 5 indicating strong agreement. The statements consist of the following:

1. Understanding how the interface works is difficult.
2. Remembering how the interface works is difficult.
3. Using the interface is difficult.
4. The interface always performs the intended action.
5. Overall, using the interface is very satisfying.
6. I will be interested to use this interface to control other physical systems.

The participant responses are tallied and used with their suggestions and comments to compare the interaction metaphors in terms of user experience. The participants report having a positive experience using the smartphone interface. Nearly all participants respond that it is neither difficult to understand nor difficult to remember how to use any of the four interface modes. However, most participants agree that the button mode is more difficult to understand at first and the most

difficult to use, that it does not always perform the participant's intended action, that it is not very satisfying, and that they would not be interested in using this interface mode again. These responses are expected, since no participant successfully completed the balancing task with the button interface. This is because the binary nature of buttons does not present a clear intuitive mapping for operating a rotating motor whose speed and direction must continuously change. This problem is solved in the gaming industry with the introduction of interface elements like knobs and joysticks. Although a better button-based mapping than the one explored in this study may exist, the pressing of buttons is severely limiting and does not bear any spatial relationship to the problem, and so it can be safely concluded that buttons cannot provide a suitable interaction metaphor for this system. In contrast to the button mode, most participants report that the tilt mode is the easiest and most intuitive to use. This is because the tilt mode effectively exploits the accelerometer onboard the smartphone to provide the user with an intuitive metaphor based on a spatial analogy which is immediately easy to understand.

Participants are also asked to select their preferred interface mode for conducting the experiment. Figure 5.8 shows the results of the responses from the participants, with 82 % of the participants preferring to use the tilt mode. Thus, in combination with the quantitative results, it is seen that the tilt mode provides the most effective, intuitive, and preferred metaphor for interacting with the ball and beam test-bed to perform the balancing task.

5.3.3 Discussion of Cyber-Physical Effects

During the interaction, participants can choose to look at the animation of the test-bed on the smartphone screen or the actual test-bed. Because of the reaction time required by the balancing task, it is generally difficult for participants to switch

Fig. 5.8 The percentage of times each interface is chosen as the preferred interface by participants
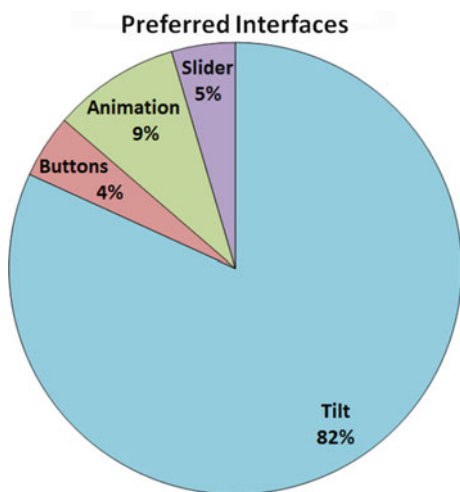
their attention between the two. Participants report that they can sense a small time delay between the response of the animated system and the physical test-bed. There will always be delays in a system such as the one presented in this chapter due to latency inherent in wireless network communication; however, under the influence of gravity, the dynamics of the ball and beam system are especially fast in comparison to the communication delay and update time of the mobile interface. This is not a significant issue in many practical applications involving systems with relatively slower dynamics and tasks which do not impose strict timing constraints [13]. In fact, noticeable delays in the animation disappear as users bring the ball to a more stable state where it can slowly move near the center of the beam. Nonetheless, most participants opt to look at the actual test-bed and use their own visual feedback as opposed to the animation. Since studies with robots show that systems that are directly teleoperated have the potential for increased performance when conducting complex tasks [4], future research must investigate how complex tasks may be performed at large distances, in which direct teleoperation is no longer accompanied by direct visibility and users must rely on only the interface. While using their own visual feedback from the test-bed, some participants report that the slider and animation interfaces, which both involve dragging fingers on the touchscreen, give essentially the same interactive experience. The tilt interface is the only interface that is unaffected by this issue, since in this case participants solely focus their attention on the actual test-bed while naturally tilting the smartphone to command the beam orientation.

Since network latency can degrade the performance and usability of the system, during the user trials efforts were made to keep latency to a minimum. Specifically, all 252 trials were conducted on the same day on a local network that was exclusively dedicated to conduct this study. Next, the size of the network packets was kept to a minimum, the distance between the mobile device and the router was held relatively constant, and algorithms [28] that degrade real-time response of the system were disabled. These strategies allowed the fluctuations in network latency to be neglected across trials and users did not experience any effects of latency fluctuations during trials.

6   Conclusions

Smart mobile devices, such as smartphones and tablets, have the potential to serve as components that may be integrated into human-in-the-loop cyber-physical systems. In these systems, they can provide users with access to immersive and intuitive interactions with complex physical processes. This chapter demonstrated that to achieve effective interactions, in terms of stability, performance, and quality of user experience, harsh demands are imposed on the mobile device that test the limits of its SSCC power. An example system was presented that illustrated how the hardware and software of mobile devices can be blended in the design of intuitive interaction metaphors that are oftentimes based on some physical analogy between user behavior and the nature of the system. Although the integration of mobile

devices using the architecture described in this chapter could serve to enhance people's interactions with physical systems, future work must further explore how to resolve issues that arise from cyber-physical effects associated with the interplay of computation, communication, and dynamics of the physical process.

Appendix: Beam Position Control Modeling and Design

To control the angular position of the beam from the laboratory station, the test-bed is modeled as a combination of an electrical and a mechanical subsystem. Figure 5.9 shows a schematic representation of the test-bed, which includes an armature-controlled DC-motor, a gearbox, and the beam [7, 20]. The governing equations for the rotation of the beam under the influence of an applied voltage to the DC-motor can be approximated as a first-order transfer function from the armature voltage $V_a(s)$ to the angular velocity of the beam, $\omega_\ell(s)$, as shown below

$$\frac{\omega_\ell(s)}{V_a(s)} = \frac{K}{\tau s + 1}, \tag{5.1}$$

where K is the dc-gain and τ is the mechanical time-constant of the test-bed dynamics. Next, the parameters K and τ are experimentally identified from the step response of the system [20] and found to be $K = 1.58\,\mathrm{rad/s/V}$ and $\tau = 0.068\,\mathrm{s}$. The first-order transfer function is combined with an integrator to obtain the angular position and converted to the following state-space model:

$$\dot{x}(t) = Ax(t) + Bu(t), \quad u(t) = V_a(t),$$

$$A \triangleq \begin{bmatrix} 0 & 1 \\ 0 & -\frac{1}{\tau} \end{bmatrix}, \quad B \triangleq \begin{bmatrix} 0 \\ \frac{K}{\tau} \end{bmatrix}, \quad x \triangleq \begin{bmatrix} \theta_\ell \\ \omega_\ell \end{bmatrix}. \tag{5.2}$$

Fig. 5.9 The model of the motorized beam
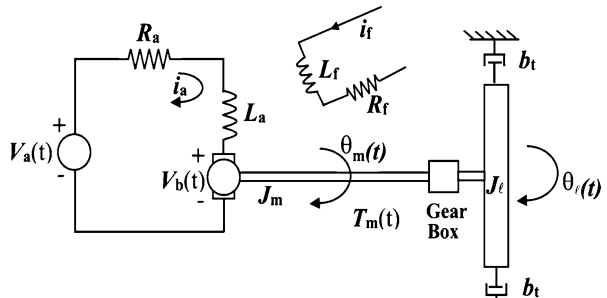
Using the second-order state-space model of the motorized beam test-bed (5.2), a linear quadratic regulator [6] is designed by selecting the control gain K_c such that the full-state feedback control law $u(t) = V_a(t) = -K_c x(t)$ minimizes the quadratic performance

$$J(K_c) = \int_0^\infty \left[x^T(t) Q x(t) + u^T(t) R u(t) \right] dt,$$

$$Q = \begin{bmatrix} 500 & 0 \\ 0 & 5 \end{bmatrix}, \quad R = 15. \tag{5.3}$$

The solution for the control gain is found to be $K_c = \begin{bmatrix} 5.7735 & 0.4765 \end{bmatrix}$. Note that a Simulink model is designed to implement the feedback controller on the desktop computer. Specifically, the control signal $u(t)$ is computed using sensor measurements and user commands received from the smartphone. To avoid the wires on the test-bed from getting entangled, in all cases, the feedback controller implementation constrains the commanded position of the beam angle to be within $\pm 20°$.

References

1. Abaid N et al (2013) Controlling a robotic fish with a smart phone. Mechatronics 23(5):491–496
2. Conti M et al (2012) Looking ahead in pervasive computing: challenges and opportunities in the era of cyber–physical convergence. Pervasive Mob Comput 8(1):2–21
3. Correa A et al (2010) Multimodal interaction with an autonomous forklift. In Proceedings of ACM/IEEE conference on human-robot interaction, pp 243–250
4. Crandall JW, Goodrich MA (2002) Characterizing efficiency of human robot interaction: a case study of shared-control teleoperation. Proc IEEE/RSJ international conference on intelligent robots and systems 2:1290–1295
5. Daponte P et al (2013) State of the art and future developments of measurement applications on smartphones. Measurement 46(9):3291–3307
6. Dorato P, Abdallah C, Cerone, V (1995) Linear quadratic control: an introduction. Prentice Hall, Upper Saddle River, NJ
7. Dorf RC, Bishop RH (2008) Modern control systems. Prentice Hall, Upper Saddle River, NJ
8. Evans AW et al (2012) Control solutions for robots using android and iOS devices. In: Proceedings of SPIE 8387, unmanned systems technology XIV
9. Fisher CJ (2010) Using an accelerometer for inclination sensing. AN-1057, Application note, Analog Devices
10. Fong T (2001) Collaborative control: a robot-centric model for vehicle teleoperation. The Robotics Institute, Carnegie Mellon University
11. Fong T, Thorpe C, Glass B (2003) PdaDriver: a handheld system for remote driving. In Proceedings of IEEE international conference on advances in robotics, Paper No. LSRO2-CONF-2003-004
12. Francese R, Passero I, Tortora G (2012) Wiimote and Kinect: gestural user interfaces add a natural third dimension to HCI. In: Proceedings of international working conference on advanced visual interfaces, pp 116–123
13. Frank JA, Kapila V (2014) Development of mobile interfaces to interact with automatic control experiments [Focus on Education]. IEEE Control Syst 34(5):78–98
14. Frank JA, Kapila V (2014) Performing difficult teleoperation tasks using dominant metaphors of interaction. In: Proceedings of ASME biennial conference on engineering systems design and analysis, Paper No. ESDA2014-20133, p V003T15A007

15. Frank JA, Gomez JAD, Kapila V (2015) Using tablets in the vision-based control of a ball and beam test-bed. In: Proceedings of international conference on informatics in control, automation, and robotics, pp 92–102
16. Frank J, Sahasrabudhe Y, Kapila V (2015) An augmented reality approach for reliable autonomous path navigation of mobile robots. In: Proceedings of Indian control conference, Paper No. WeBT2.2
17. Gerling KM, Schulte FP, Masuch M (2011) Designing and evaluating digital games for frail elderly persons. In: Proceedings of international conferences on advances in computer entertainment technology, p 62
18. Guo C, Sharlin E (2008) Exploring the use of tangible user interfaces for human-robot interaction: a comparative study. In Proceedings of SIGCHI conference on human factors in computing systems, pp 121–130
19. Hochberg Y, Tamhane AC (1987) Multiple comparison procedures. Wiley, New York, pp 96–98
20. Kapila V, Lee S (2008) Automatic control laboratory experiment manual, 2nd edn. Polytechnic Institute of New York University
21. Khaitan SK, McCalley JD (2014) Design techniques and applications of cyberphysical systems: a survey. IEEE Syst J 9(2):350–365
22. Khan WZ et al (2013) Mobile phone sensing systems: a survey. IEEE Commun Surv Tutorials 15(1):402–427
23. Krofitsch C et al (2013) Smartphone driven control of robots for education and research. In: Proceedings of IEEE international conference on robotics, biomimetics, and intelligent computational systems, pp 148–154
24. Kuparinen L, Silvennoinen J, Isomäki H (2013) Introducing usability heuristics for mobile map applications. In: Proceedings of international cartographic conference
25. Lane N et al (2010) A survey of mobile phone sensing. IEEE Commun Mag 48(9):140–150
26. Lopez DA, Frank JA, Kapila V (2014) Comparing interface elements on a tablet for intuitive teleoperation of a mobile manipulator. In: Proceedings of ASME biennial conference on engineering systems design and analysis, Paper No. ESDA2014-20126, p V003T15A005
27. Milgram P (1993) Applications of augmented reality for human-robot communication. In: Proceedings of IEEE/RSJ international conference on intelligent robots and systems, pp 1467–1472
28. Mogul JC, Minshall G (2001) Rethinking the TCP Nagle algorithm. ACM SIGCOMM Comput Commun Rev 31(1):6–20
29. Nielsen (2010) The state of mobile apps. The Nielsen Company. http://www.nielsen.com/content/dam/corporate/us/en/reports-downloads/2010-Reports/Nielsen-State-of-Mobile-Apps-Whitepaper.pdf. Cited 30 Jan 2014
30. Nielsen CW, Bruemmer DJ (2007) Hiding the system from the user: moving from complex mental models to elegant metaphors. In: Proceedings of IEEE international symposium robot and human interactive communication. Engineering systems design and analysis, pp 756–761
31. Nielsen J, Budiu R (2012) Mobile Usability. New Riders, Berkeley
32. Schirner G et al (2013) The future of human-in-the-loop cyber-physical systems. Computer 46(1):36–45.
33. Sheridan TB (1992) Telerobotics, automation, and human supervisory control. MIT Press, Cambridge
34. Walker AM, Miller DP (2012) Tele-operated robot control using attitude aware smartphones. In: Proceedings of ACM/IEEE conference on human-robot interaction, pp 269–270
35. Walters ML et al (2005) Practical and methodological challenges in designing and conducting human-robot interaction studies. Proc AISB 5:110–119
36. Wasserman AI (2010) Software engineering issues for mobile application development. In: Proceedings of FSE/SDP workshop future of software engineering research, pp 397–400
37. White J et al (2010) R&D challenges and solutions for mobile cyber-physical applications and supporting internet services. J Internet Serv Appl 1(1):45–56
38. Williams DH (2003) PDA robotics: using your personal digital assistant to control your robot. McGraw-Hill, New York

Chapter 6
Force/Tactile Sensors Based on Optoelectronic Technology for Manipulation and Physical Human–Robot Interaction

Andrea Cirillo, Pasquale Cirillo, Giuseppe De Maria, Ciro Natale, and Salvatore Pirozzi

Abstract Design and realization of autonomous robotic platforms require crucial information on their surroundings especially when robots should interact with the environment and humans. In many cases, perception is necessary to correctly accomplish tasks in a dynamic environment where a model is hard to obtain. Grasping and fine manipulation of objects with different shapes and surface characteristics as well as detection of contacts between the robot and the environment are easily enabled by tactile sensing. The sense of touch represents the most natural way to obtain relevant information during an interaction task, such as parameters like surface friction, exchanged forces and torques, object shape. This chapter provides an overview of the authors' work on force/tactile sensors development. By exploiting optoelectronic technology, the authors designed and realized two tactile sensors that can be used to execute both fine manipulation of objects and safe interaction tasks with humans. The chapter describes both sensors in detail and provides an experimental validation of their capabilities.

Keywords Force/torque sensor • Tactile sensor • Distributed sensor • Optoelectronic sensor • Human–robot interaction • Manipulation • Grasping • Slipping detection • Conformable sensor • Slipping avoidance • Friction estimation

1   Introduction

Robotic technologies have improved efficiency and reduced costs of repetitive, well-defined manufacturing tasks. Recently, there is a great interest in designing robots that can work in less-structured environments by collecting information about

A. Cirillo • P. Cirillo • G. De Maria • C. Natale (✉) • S. Pirozzi
Dipartimento di Ingegneria Industriale e dell'Informazione, Seconda Università degli Studi di Napoli, via Roma 29, 81031 Aversa, Italy
e-mail: andrea.cirillo@unina2.it; pasquale.cirillo@unina2.it; giuseppe.demaria@unina2.it; ciro.natale@unina2.it; salvatore.pirozzi@unina2.it

their surroundings to take appropriate decisions [48]. This would allow them to work in areas close to humans and perform more complicated and dynamic tasks (e.g., providing basic services to elderly people or undertaking dangerous rescue missions) [37]. Autonomous or cooperative tasks require the robots to safely interact with the environment, i.e., without producing damages to themselves, to humans, and to objects.

In this scenario, sensing becomes fundamental. Tactile sensing is particularly important since many tasks require the robot to recognize unintentional collisions or to make intentional physical contact with humans or objects. Highly functional tactile sensing is required to improve safety and effectiveness of current robotic technology [24]. For humans, the skin is a remarkable organ and the sense of touch is fundamental to grasp and manipulate objects avoiding slippage, or to blindly operate in a dynamic environment. It consists of an integrated, stretchable network of sensors that provides information about tactile, proximity, and thermal stimuli to the brain, allowing us to operate within our environment safely and effectively. An artificial tactile sensor, mimicking the human touch, should possess the capability to measure both dynamic and geometric quantities, i.e., contact forces and torques as well as spatial and geometrical information about the surface of the touched objects. Each of these quantities may be measured either as an average quantity for some parts of the robot or as a spatially resolved, distributed quantity across a contact area [54]. Many essential aspects of life are mediated by the multi-functional tactile sensing capabilities of skin, including contact force sensing for grasp control, object manipulation and orientation determination, tensile strain monitoring for proprioception (essential for simple movements such as standing or walking), shear force sensing for friction determination, vibration detection, slip detection, and texture determination.

As Mallory et al. said in [25], the idea of developing force/tactile sensors that mimic the capabilities of the human skin has been used for the first time several years ago also in science fiction and movies, e.g., Six Million Dollar Man series, Star Wars series, and Terminator movie series. Since 1990s, scientists began to create low-cost, conformable, and useful tactile sensors for normal and shear forces estimation [12, 32], but significant progress in the development of tactile sensors has been achieved just in recent years, when robotics researchers tried to improve the cognitive and grasp capabilities of new robots by providing them with both compact tactile sensors for robotic hands and distributed sensor systems for arms and body. A recent and comprehensive review on features that tactile sensors should possess and technologies used to realize them can be found in [15]. Scientific literature reports a lot of works related to design and development of small and compact tactile sensors as well as of distributed prototypes that use different technologies, e.g., resistive [1, 4, 23, 30, 43, 53, 56, 58, 59], piezoelectric [35, 42], capacitive [5, 8, 13, 22, 44, 45, 50–52], magnetic [7], and optoelectronic [26, 27, 33, 40, 46]. While the use of tactile sensors for contact point detection and pressure estimation is a diffuse practice, the development of force/tactile sensors able to estimate the contact wrench as well as the design of distributed solutions able to estimate both magnitude and direction of the applied forces remains still an open challenge for

robotics researchers. Other examples of sensors that can detect contact pressure have been presented in [19, 20, 29, 36, 47, 55, 57]. To the best of the authors' knowledge, the optoelectronic technology is not widely used in the design of tactile sensors. Just few solutions, such as the one proposed in [40], exploit optoelectronic technology to obtain a six-axis force sensor. A vertical cavity surface emitting laser (VCSEL) is used in [31, 41]. In [41], the VCSEL, combined with a photodiode, is used to develop a flexible sensor able to estimate the shear stresses applied to its surface. Differently, in [3] an optic subsystem is introduced only to provide a reliable and high speed communication channel for the sensor data transmission and not for the force measurement. None of the sensors cited so far is able to estimate the force vector in a large contact area and only other few different non-optoelectronic solutions are able to estimate the three components of the force vector applied to a small contact surface by an external colliding object. In [60, 61] a flexible tactile sensor array for an anthropomorphic artificial hand with the capability of measuring both normal and shear force distributions, using quantum tunneling composite (QTC) as a base material, is presented. It consists of four fan-shaped electrodes in a cell that decomposes the contact force into normal and shear components. However, as discussed by Amarasinghe et al. in [2], QTC technology suffers from different drawbacks that imply to use it or as a simple low-cost contact switch, or with additional complex mechanical structures to realize force sensors. Moreover, QTC also takes considerable time to return to its original state after deformation. Tao Liu et al. [38] designed a 3-D tactile sensor by integrating four sensing cells, each composed of a pressure-sensitive electric conductive rubber (PSECR) and a fan-shaped pectinate circuit. Both solutions use measurements of four sensing cells to estimate the normal force component and the shear force components through a simple empiric relationship that relates the force components to the sensor output, i.e., voltages.

The solutions cited above involve large volumes, weights, and costs that complicate the integration especially into robotic hands.

In this chapter the authors present their own solution based on the use of optoelectronic technology aimed to overcome most of the problems encountered in the works cited above, mainly: difficulty of integration into small spaces, high costs, repeatability, and complex conditioning electronics. The idea behind the sensor design has been exploited for the realization of a six-axis force/tactile sensor and of a smaller three-axis sensor as well. The first one, thought for improving grasping and object manipulation capabilities of robotic hands, is able to measure the six components of the force and torque vectors applied to it and to provide a spatial and geometrical information about the contact with a stiff external object. The second one represents a smaller and simplified version of the previous sensor and it has been used like the tactile element for the development of a spatially distributed force/tactile sensor prototype. The developed sensor prototypes have been characterized and tested in typical robotics applications of manipulation, like slipping detection and human–robot interaction, i.e., manual guidance, collision detection and reaction.

Concepts and ideas used for design and development of the sensors presented in this chapter have been already presented by the authors in [9, 10, 16, 17], in [18] where the six-axis sensor has been used for slipping detection and avoidance, and in [11] where the distributed sensor has been used for safe physical human–robot interaction. This chapter is aimed at showing how optoelectronic technology has been exploited to realize reliable and efficient force/tactile sensors for robotics applications.

2   Sensor Concept

To realize the sensor with all the capabilities stated in the introduction, a deformable elastic layer is positioned above a matrix of sensible points (taxels) so as to transduce the force and/or torque vectors into deformations, which are measured by the sensible points, as explained in the following subsection. Furthermore, the signals provided by the taxels, which are spatially distributed below the deformable layer, constitute a spatially distributed information that will also allow to estimate the size and orientation of the contact area between the external surface of the sensor and the objects in contact with it. The contemporary knowledge of all these information is essential for using the sensor in robotics applications where objects of different size and dimension have to be manipulated by robotic hands.

The sensor concept is based on the use of couples of infrared LED-phototransistor and of a deformable elastic layer positioned above the optoelectronic devices (see Fig. 6.1). The optoelectronic components are organized in a matrix structure. For each couple, the LED illuminates the reflecting surface that coincides with the bottom facet of the deformable layer. The deformable layer transduces an external force and/or torque into a deformation of its bottom facet. An external force applied to the deformable layer produces local variations of the deformable material and the couples of optical devices measure such deformations in a discrete number of points. These deformations produce a variation of the reflected light

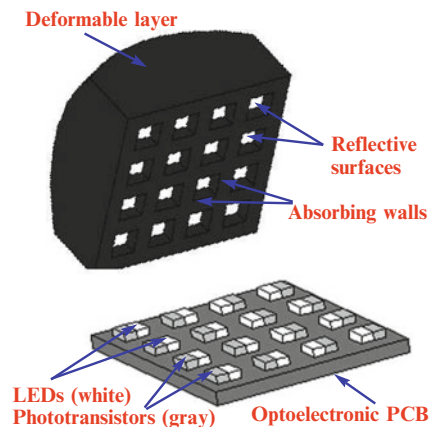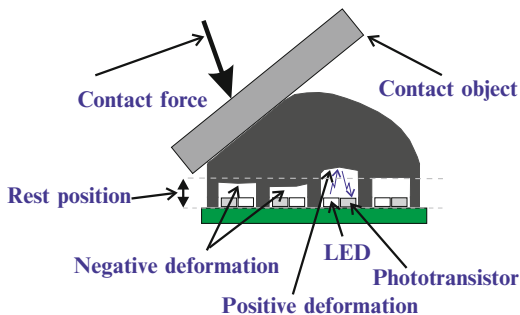**Fig. 6.1** 3D CAD model of the tactile sensor

Fig. 6.2 Schematization of
the working principle

intensity and, accordingly, of the photocurrent flowing into the photodetector. In
particular, the deformations can be positive or negative, i.e., the photocurrent can
locally increase or decrease (see Fig. 6.2), depending on amplitude of tangential and
normal force components, as well as on torque components. The authors reported
in [14] a FE analysis demonstrating that the described relationship can be used to
actually reconstruct the external force and/or torque components by measuring the
elastic layer deformations in a discrete number of points.

3   The Force/Tactile Sensor

According to the working principle described in Sect. 2 different sensor prototypes
have been developed. In the following, a tactile sensor suitable to be installed on
the fingertips of anthropomorphic hands is described. Given its small dimension
and weight and its capabilities, the sensor can be used to confer tactile sensing to
robotic hands allowing them to grasp and manipulate objects with different shapes.

3.1   Realization of the Prototype

The realized tactile sensor is able to measure the six components of the force and
torque vectors applied to its deformable surface. It consists of 16 taxels organized
in a 4×4 matrix. The deformable layer is made of black silicone in order to
avoid cross-talk problems between taxels and ambient light disturbance, since the
black color guarantees the maximum absorption at every wavelengths. Only the
surface which faces each taxel is white to increase sensor sensitivity, ensuring
maximum reflection at every wavelengths (see Fig. 6.3b). According to FE analysis
results, the aspect ratio of the black walls between taxels has been selected in
order to reduce the horizontal deformations with respect to the vertical ones. In
particular, the thickness of the black walls is 1 mm, while the area of the white
reflecting surfaces is 1.6×1.6 mm, resulting into a total size of the deformable
layer of 11.4×11.4 mm. The height of the reflecting surfaces from the base of
the deformable layer is 1.5 mm thus ensuring the necessary aspect ratio. The top of
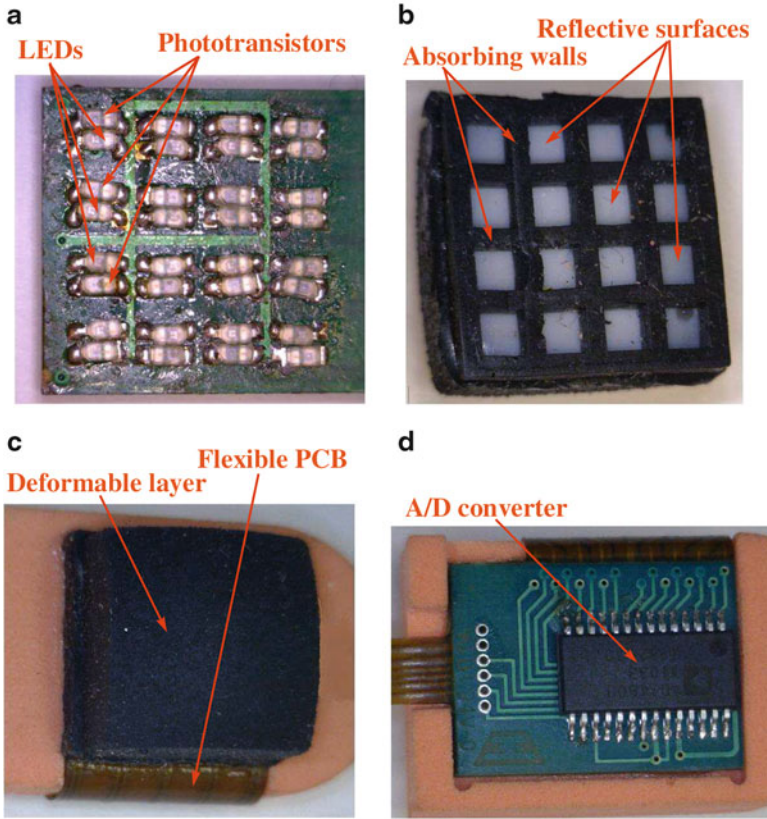
Fig. 6.3 Pictures of the tactile sensor: (**a**) electronic layer, (**b**) bottom view of deformable layer, (**c**) top view of assembled sensor, (**d**) bottom view of assembled sensor

the deformable layer is a section of a sphere with a radius of 11.4 mm. According to the numerical simulations reported in [14], the expected measurement range of the sensor prototype is [0, 4] N. The maximum force level can be adapted by changing the hardness of the deformable layer. The maximum measurable force is limited by the maximum vertical deformation of the reflecting surface of each taxel, so the former can be changed by acting on the deformable layer geometry. A linear relation between the Shore hardness and the logarithm of the Young's modulus has been derived in [49] for elastomeric materials. Using this relation, the maximum predictable force level, with the current geometry, goes from 2 to 40 N by changing the hardness of the deformable layer from 4 to 60 A.

Recalling the theory on LEDs and photodetectors radiation pattern [34], it is possible to consider the photocurrent (and thus the received radiant flux by the photodetector) proportional to the radiant intensity pattern of the LED and responsivity pattern of the photodetector, evaluated at the emitting and receiving

angle, respectively. Moreover, the photocurrent is also inversely proportional to
the squared distance between the two devices and the reflecting surface. As a
consequence, the optoelectronic components should have very large angles of
view in order to minimize the effects of LED radiation pattern and photodetector
responsivity pattern on the photocurrent, thus leaving only the dependence with
the distance. Considering these aspects, for the realization of the prototype the
following optoelectronic components manufactured by OSRAM (Fig. 6.3a) have
been selected. The LED (code SFH4080) is an infrared emitter with a typical
peak wavelength of 880 nm, while the detector is a silicon NPN phototransistor
(code SFH3010) with a maximum peak sensitivity at 860 nm wavelength. Both
components have an angles of view of $\pm 80\,°$. The conditioning electronics is only
constituted by simple resistors without amplification and/or filtering stages. For this
reason, the collector current of the phototransistors, depending on the received light
intensity, is easily translated into a voltage signal via a suitable resistor and then
directly digitized with an analog-to-digital converter chip (code AD7490, Fig. 6.3d).
The voltage supply for all components is 3.3 V. Each LED is used with a forward
current less than 5 mA, with a maximum power consumption less than 200 mW for
the whole tactile sensor. This low power consumption and the local digitalization,
along with its compact size (Fig. 6.3c), make this sensor particularly suitable for
integration into anthropomorphic robotic hands.

3.2   *Characterization as Force/Tactile Sensor*

The objective of this section is to show the characteristics of the presented tactile
sensor and a calibration procedure necessary to use it as a 6-axis force/torque sensor.
The characterization of the sensor has been made under the assumption that the
contact surface can be approximated by a plane with a high stiffness with respect
to the deformable layer. The hypothesis holds if the external object has a curvature
radius larger than that of the deformable layer. This happens for a large number of
objects used in common manipulation and grasping tasks. Taking into account the
hardness of the silicone used to realize the sensor prototype, also the hypothesis of
high stiffness for the contact surface can be considered true for most of the daily
use objects. Figure 6.4 shows a sketch of the sensor, where the position of each
cell with respect to the reference frame is indicated. The position of the k-th taxel
can be identified with the coordinates of the taxel center position (x_k, y_k). When the
silicone layer is undeformed, a certain amount of light emitted by the LED of each
taxel is reflected from the white surface and it is received by the phototransistor,
generating an initial voltage value on the collector. When an external force and/or
torque is applied to the sensor, the distance of the reflecting surface of each cell
from the corresponding taxel on the electronic layer can be subjected to a positive
or a negative variation. This implies changes of the reflected light intensity and,
accordingly, of the voltages on the phototransistor collectors. Denoting with v_k the
voltage variation of the k-th taxel, $v_k > 0$ denotes an increasing distance (and then a

Fig. 6.4 The tactile sensor
reference frame
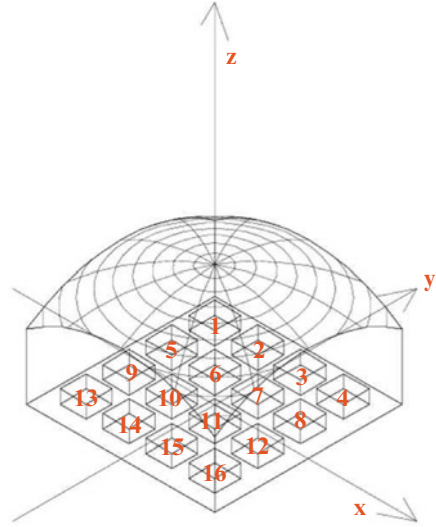
Fig. 6.5 Calibration system
for the tactile sensor as
force/torque sensor
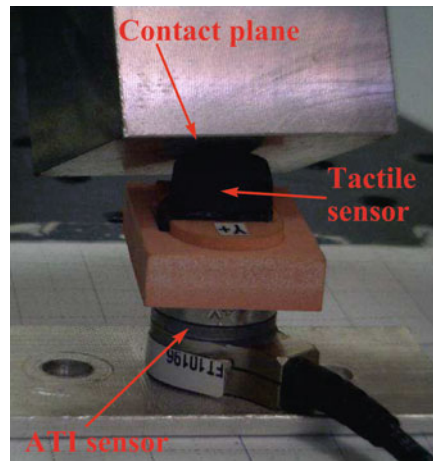
decreasing photocurrent), while $v_k < 0$ denotes a decreasing distance (and then an
increasing photocurrent) between the reflecting surface and the electronic layer. To
calibrate the prototype as a six-axis force/torque sensor, the proposed approach is
based on the use of a neural network to model the relationship between the applied
forces and torques and the phototransistors measurements. The sensor has been
mounted on a six-axis load cell used as reference sensor. The model used is the
FTD-Nano-17 manufactured by ATI, with a measurement range equal to ± 12 N
and ± 17 N for horizontal and vertical force components, respectively, while the
measurement range for all torque components is equal to ± 120 Nmm. The reference
axes, reported in Fig. 6.4, are positioned on the plane that separates the tactile sensor
from the reference sensor. Figure 6.5 shows the calibration system used to collect

the data for neural network training. An operator carried out various experiments, using a stiff plane, applying different external forces and torques and simultaneously acquiring all the voltage variations on the phototransistors and all the forces and torques components measured by the reference load cell. The data, acquired at a sample rate of 100 Hz, have been organized in a training set and a validation set to be used as input data (voltage variations) and target data (forces and torques components) of the neural network. A testing set has been prepared using data from experiments other than those used to collect the training and the validation sets, in order to assess the trained network. A standard two-layer feed-forward neural network $\mathbf{f}_{NN}$, trained with the Levenberg–Marquardt method, has been used to fit training data

$$\mathbf{H} = \mathbf{f}_{NN}(\mathbf{V}), \tag{6.1}$$

where $\mathbf{H} = [F_x, F_y, F_z, T_x, T_y, T_z]^T$ is the output vector (expressed in the sensor frame fixed to the base of the sensor with the origin in the center of the spherical deformable layer, the z axis orthogonal to the electronic layer and pointing outwards, the x and y axes placed as indicated in Fig. 6.4), F_i and T_i with $i = x, y, z$ are force and torque components with respect to the reference frame, and $\mathbf{V} = [v_1, v_2, \ldots, v_{16}]^T$ is the input vector containing the voltage variations of the taxels. The network is constituted by 24 neurons for the hidden layer and 6 neurons for the output layer. The trained network testing results are reported in Fig. 6.6a for the force components and in Fig. 6.6b for the torque. The estimation is good for all force and torque components, especially when force and torque values are high enough. In fact, in the performance analysis it must be also taken into account that the training data are very noisy when the measured values are small compared to the full scale of the reference sensor. It is important to underline that these performances have to be evaluated considering that, compared to commercially available sensors, the proposed one is more compact, low cost, low power consumption, provided with a digital interface, and the deformable layer guarantees good adaptability and stability during grasping and manipulation applications.

The same taxel signals, used to estimate forces and torques on the contact point, represent also a distributed information, commonly called pressure map or tactile image. The tactile image can be used to estimate the center of pressure (CoP) location at the base of the deformable layer, whose coordinates with respect to the sensor frame can be estimated through a centroid-like method, i.e.,

$$x_H = \frac{\sum_{k=1}^{16} \alpha_k G(v_k)}{\sum_{k=1}^{16} G(v_k)}, \ y_H = \frac{\sum_{k=1}^{16} \beta_k G(v_k)}{\sum_{k=1}^{16} G(v_k)}, \tag{6.2}$$

where

$$G(v_k) = \begin{cases} v_k & \text{if } v_k \leq 0 \\ 0 & \text{if } v_k > 0 \end{cases} \quad k = 1, \ldots, 16. \tag{6.3}$$
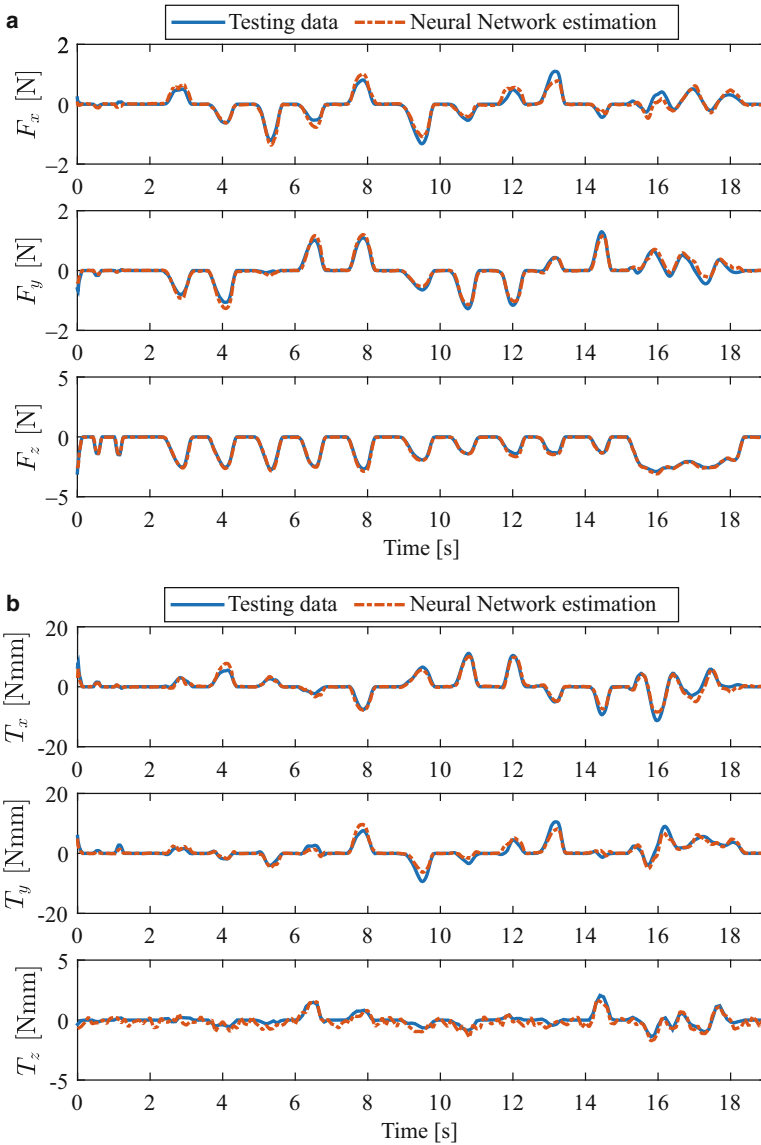
Fig. 6.6 Neural network testing: (**a**) force component estimation, (**b**) torque component estimation

Figure 6.7a reports an example of a linearly interpolated pressure map with the estimated CoP.

The CoP coordinates obtained from tactile data are fundamental for the estimation of the contact plane orientation with respect to the sensor frame. For a generic contact plane, it is possible to fix a contact frame as reported in Fig. 6.7b,
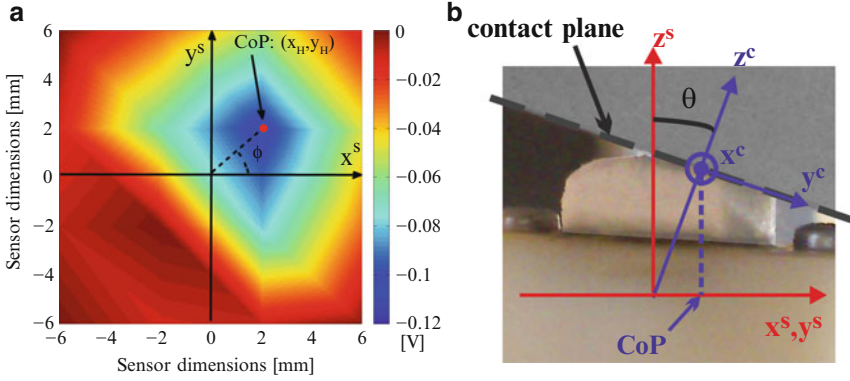
Fig. 6.7 The tactile sensor: (**a**) example of pressure map with the CoP and definition of the ϕ angle; (**b**) definition of the θ angle with respect to the contact frame

whose orientation with respect to the sensor frame is defined by the angles ϕ and θ. A simplified version of the relationship between CoP coordinates and contact plane orientation presented in [17] can be obtained by neglecting the deformations parallel to the contact plane

$$x_H = Rc_\phi s_\theta, \quad y_H = Rs_\phi s_\theta, \tag{6.4}$$

where R is the radius of the hemispherical deformable layer and the symbols c_ϕ and s_ϕ denote $\cos\phi$ and $\sin\phi$, respectively. By inverting eq. (6.4) and given the CoP coordinates computed as in eq. (6.2), the angles θ and ϕ can be estimated. The knowledge of the orientation angles of the contact plane allows to rotate the force vector, expressed in the sensor frame, in order to estimate the force vector expressed in the contact frame, namely

$$\mathbf{f}^c = \begin{pmatrix} c_\theta c_\phi & c_\theta s_\phi & s_\theta \\ -s_\phi & c_\phi & 0 \\ -s_\theta c_\phi & -s_\theta s_\phi & c_\theta \end{pmatrix} \mathbf{f}^s, \tag{6.5}$$

where $\mathbf{f}^s = (f_x^s \ f_y^s \ f_z^s)^T$ is the force vector in the sensor frame and $\mathbf{f}^c = (f_x^c \ f_y^c \ f_z^c)^T$ is the force vector in the contact frame. From $\mathbf{f}^c$ it is straightforward to compute normal and tangential components of the contact force as

$$f_n = |f_z^c| \tag{6.6}$$

$$f_t = \sqrt{f_x^{c2} + f_y^{c2}}. \tag{6.7}$$

4   The Distributed Force/Tactile Sensor

The tactile sensor described in Sect. 3 represents a reliable solution to improve grasp and manipulation capabilities of robotic hands given the possibility to measure both force and torque vectors that act on it. However, such sensor is not suitable to be scaled in such a way it can be installed on a large robot surface since it is not conformable. In general, a distributed force/tactile sensor should possess the following features:

- capability to provide a direct measure of the pressure map and of the contact force vector so as to avoid complex model-based estimation algorithms, even in the case of multiple contacts
- high modularity and scalability
- light-weight, low power consumption, and low-cost, since it potentially could be used to cover the entire surface of a robot
- adaptability to different spatial resolutions requested by the application site
- easy of manufacturing with a low number of wires
- adaptability to arbitrarily curved surfaces (conformability)

For the design of a distributed force/tactile sensor, the features to be taken into account are different from those of a fingertip tactile sensor, since the main objective of a distributed sensor is not the fine manipulation but the safe human–robot interaction. In particular, some characteristics such as spatial resolution, accuracy of the force estimation, and sensitivity can be relaxed in favor of additional features such as modularity, possibility to cover large areas with limited costs and power consumption, capability to discriminate multiple contact areas with the corresponding forces, and ease of integration in different parts of the robot.

4.1   The Sensing Module

According to the guidelines defined in the previous discussion, the authors proposed a distributed sensor based on the same idea behind the tactile sensor. The distributed force/tactile sensor is obtained by interconnecting several sensing modules, each of them capable to measure the force vector acting on it. Each sensing module consists of four taxels organized in a 2×2 matrix. As before, a single taxel consists of an optical emitter/receiver couple spectrally matched and a deformable elastic layer is positioned above the 4 optoelectronic couples. Differently from a fingertip tactile sensor, a distributed force/tactile sensor requires a higher full scale. To obtain these characteristics, a silicone with a higher hardness has been chosen. The selected one is the MM928, provided by ACC Silicones Europe, with a Shore hardness of 28 A and a cure time of 24 h at room temperature. As in Sect. 3, the aspect ratio of the black walls between taxels has been selected in order to reducehorizontal
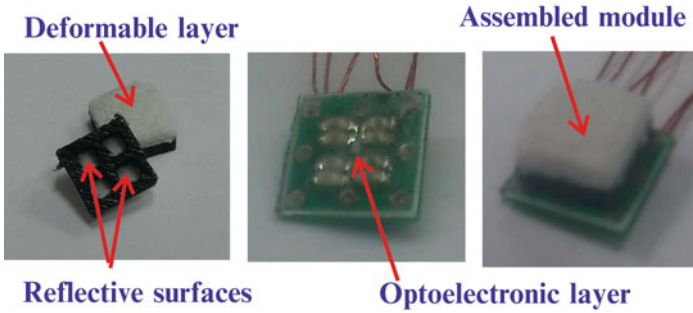
Fig. 6.8 Pictures of the sensing element

deformations with respect to vertical ones, according to the finite element analysis reported in [14]. In particular, for the distributed prototype, the thickness of the black walls is 0.8 mm, while the extension of the white reflecting surfaces is 1.8 ×1.8 mm, which results in a total size of the deformable layer of 6×6 mm. The height of the reflective surfaces from the electronic layer is 1.6 mm. The top of the deformable layer is a section of a sphere with 7 mm radius. The deformable layer is bonded on the electronic layer (of size 6.4 ×6.4 mm) by cyanoacrylate glue. Figure 6.8 shows some pictures of the sensing module components and an assembled module. After a suitable calibration process, which is deeply described in Sect. 4.4, the sensing module is able to measure the three components of the force applied to it with a good accuracy and a low hysteresis error. In [9] the authors report a study of the sensor and its characterization in terms of accuracy, hysteresis and response time, which are briefly summarized below:

- sensitivity: $\simeq 0.2\,\text{N}$
- repeatability error: $\simeq 6\,\%$
- hysteresis error: $\simeq 10\,\%$
- response time: $\simeq 0.001\,\text{s}$

An important remark concerns the hysteresis error. Its limited value allows to use the sensor information without introducing further compensation algorithms differently from other solutions based on soft foam materials [28].

Several sensing modules can be interconnected together in order to realize a distributed force/tactile sensor. The adopted choice allows to estimate the force vectors wherever applied to the whole distributed sensor by guaranteeing all discussed features. The selected sensor architecture, differently from the tactile sensor described in Sect. 3, results scalable enough to be applied to robot surfaces such as torso, legs, and arms: its spatial resolution can be properly adapted on the basis of the robot body part to cover by simply changing the distance between two adjacent sensing modules.

4.2   From the Sensing Module to the Distributed Sensor

The distributed sensor prototype described in this chapter consists of 36 modules, organized as a 6×6 matrix, for a total of 144 taxels. To ensure the scalability and the modularity of the system, a "scanning control" strategy, based on the same idea reported in [47], has been adopted to realize the module interrogation by using an MCU. The basic idea is to connect the sensing modules in groups which share 4 ADC channels, and to use the digital I/O of the MCU to switch on and off the sensing modules, with a cyclic pattern, by ensuring that in each time instant, for each group, only one taxel is turned on, while all the others, which share the same ADC, are turned off. This control logic is based on the fact that the switched off photodetectors behave as an open circuit that does not influence the A/D conversion of the switched on photodetector voltage. Differently from [47], the sensing modules can be directly driven by the MCU digital I/O lines, without using an external power supply, since each LED works with a forward current of about 1 mA and the voltage supply for all components is the 3.3 V, available from the MCU. Since different groups use different A/D channels, sensing modules belonging to different groups can share the same digital I/O line as power supply, by reducing also the number of digital I/O lines necessary to switch on and off the sensing modules during the interrogation. The described scanning strategy provides several advantages: a reduction of the whole sensor power consumption, since the number of modules simultaneously turned on is limited; a reduced number of ADC channels required to acquire the data and a simplification of the wiring. By generalizing the adopted interrogation technique, a total of n sensing modules (corresponding to $4n$ taxels) can be organized in m groups, each one constituted by p sensing modules. The sensing modules of each group share 4 A/D channels, hence the number of external wires needed to interrogate a sensor patch is equal to $4m + p$ (plus one for the ground). As a consequence, to minimize the number of wires needed for a sensor patch, the following constrained optimization problem can be solved:

$$\min_{m,p} \ (4m + p)$$

$$\text{subject to: } mp = n, \quad m, n, p \in \mathcal{N}^+. \tag{6.8}$$

By solving the optimization problem (6.8) for the 144 taxels sensor, the parameters that allow the use of the minimum number of wires are $m = 3$ groups (corresponding to $4m$ A/D channels) and $p = 12$ digital I/O lines, for a total of 24 wires plus one for the ground.

Summarizing, the 144 total taxels, which constitute 36 sensing modules organized in three groups, are interrogated by using 12 ADC channels (4 ADC channels shared by each group) and 12 digital I/O lines used to implement the scanning strategy (see the schematic diagram shown in Fig. 6.9). For applications where large surfaces have to be covered, the distributed force/tactile sensor presented in this chapter presents very attractive properties from the power consumption point of
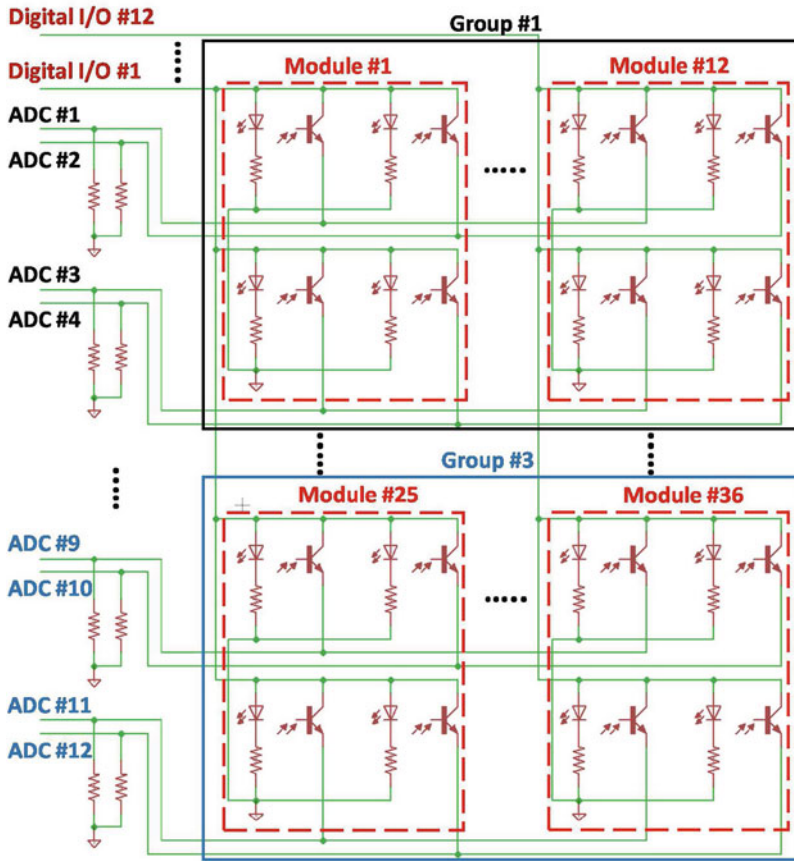
Fig. 6.9 Electronic scheme of the interconnections between several sensing modules of a distributed sensor and a microcontroller

view. Each taxel requires a voltage supply equal to 3.3 V with a current of about 1 mA, for an instantaneous power consumption of 3.3 mW. Since no additional integrated circuits (ICs) are necessary, with just a few watts of power, thousands of taxels can be driven at the same time. Generalizing, k taxels require a power consumption equal to $k \cdot 3.3$ mW. For a distributed sensor constituted by 144 taxels, a total instantaneous power consumption of 475, 2 mW would be needed if all taxels were always switched on. In this case, the power consumption would already be quite limited, but the interrogation technique described above allows a further power saving. In particular, at each time instant, only one sensing module is switched on for each group, corresponding to 4 m taxels. With the optimal number of groups $m = 3$, only 12 taxels are switched on at the same time, with a total instantaneous power consumption of 39.6 mW, resulting in a reduction of one order of magnitude compared to the previous case. The only limitation can be the

minimum sampling frequency necessary to interrogate the whole distributed sensor. For all the 144 taxels of the proposed sensor, with the selected MCU, i.e., an ARM Cortex M4 STM32F303, a sampling frequency of 150 Hz was obtained. Therefore, the proposed solution is very attractive for battery-powered robotic systems. The interrogation circuitry described so far has the great advantage that the sensor can work also if not all the sensing modules are actually connected or some connected modules are broken, since they appear only as open circuits. This improves the modularity of the proposed solution, because the number of the actually installed and/or working modules can be easily detected through a proper initialization phase.

The adopted smart scanning control strategy and the low power consumption of a sensing module made possible the development of the distributed sensor prototype with flexible PCB technology in order to enhance the sensor conformability. The scanning strategy allows a substantial reduction of the number of wires with respect to the taxel number, which makes possible the use of a flexible PCB with a limited number of layers corresponding to a reduced thickness, thus guaranteeing high conformability of the optoelectronic board and low production cost.

The design of the flexible PCB affects the maximum achievable flexibility of the distributed sensor. First of all, the installation of the electronic components on the flexible PCB reduces the flexibility, depending on number and dimensions of the components. Moreover, the flexibility depends also on the number of layers necessary for the wiring, thus a proper routing of the PCB should be carried out. The sensing modules are only constituted by the optoelectronic components (SFH4080 and SFH3010), that have a SmartLED package 0603 (with dimensions $1.3 \times 0.8 \times 0.65$ mm), and additional resistors to drive the LEDs (a resistor for each LED), with package 0402 (with dimensions $1 \times 0.5 \times 0.32$ mm). By using the specific scanning strategy, each group of sensing modules can share the resistors in series with the PTs. With this choice, the number of resistors needed to convert the photocurrents into voltages acquired by the A/D channels is reduced from the number of PTs to the number of A/D channels used during the scanning. Furthermore, the resistors can be mounted directly near the A/D channels, avoiding to add components on the conformable part of the PCB. By adopting the presented scanning strategy, the routing of a whole conformable force/tactile sensor patch can be completed by using a flexible PCB with only four layers. Since no additional ICs with cumbersome package are used for the conditioning electronics, only three types of components have to be mounted on the flexible PCB for each taxel, and they are small enough to maintain a high flexibility. The layout of the obtained PCB is reported in Fig. 6.10, including the dimensions. The active surface of the sensor patch (hosting the sensing elements) is about 47×47 mm^2, while the 25 wires, needed to interrogate the patch, are routed to a standard connector positioned on the left side. Figure 6.11 reports a picture of the realized PCB, where the high flexibility is highlighted. The solution,

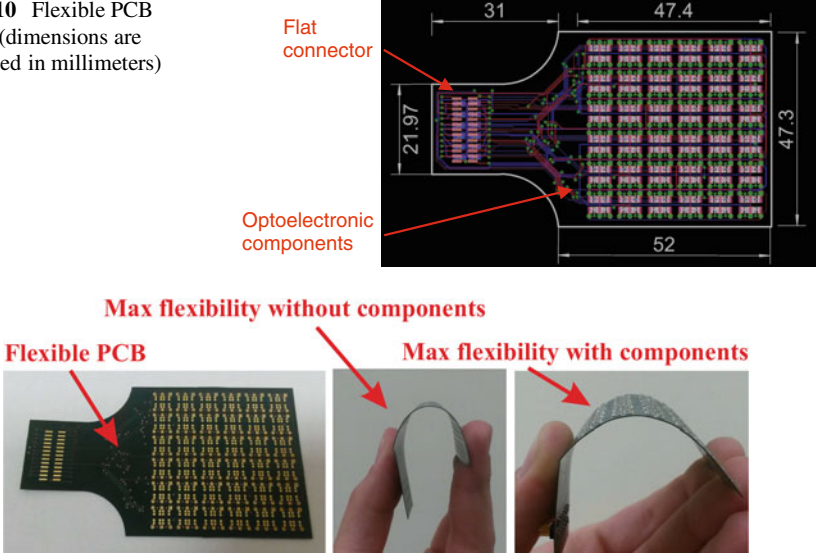**Fig. 6.10** Flexible PCB design (dimensions are expressed in millimeters)

Fig. 6.11 Pictures of the flexible PCB before (*left* and *center*) and after (*right*) soldering of optoelectronic components

after components soldering, maintains a high flexibility that allows the sensor patch to be conformed to a surface with minimum curvature radius of about 3 cm that is sufficient for covering robot surfaces such as arms, legs, and torso.

As reported in the previous section, the deformable layer of each sensing module has been bonded on the electronic layer using a cyanoacrylate glue. In order to improve the bonding reliability, to increase the loading cycles and in order to cover the whole sensor matrix with a single deformable layer, all the sensing elements are connected together by an additional silicone molding. For a reliable contact force estimation, it is necessary to have a negligible mechanical coupling between adjacent sensing elements. To minimize such a mechanical coupling, a FE analysis has been conducted (originally presented in [10]) to verify that the use of a silicone with shore hardness of 6 A (4–5 times more soft with respect to the one used to realize the deformable layer) would not transmit significant stress from one cell to another. The silicone rubber behavior was modeled with the Mooney–Rivlin constitutive law

$$\sigma = 2 \left(\lambda - 1/\lambda^2\right) \left(\alpha_1 + \alpha_2/\lambda\right), \tag{6.9}$$

where λ is the elongation ratio. The model parameters α_1 and α_2 have been evaluated based on the considerations reported in [14]. They have been chosen as $\alpha_1 = 3.96 \times 10^{-2}$ and $\alpha_2 = -3.37 \times 10^{-4}$ for the silicone with shore hardness 6 A, while as $\alpha_1 = 0.16$ and $\alpha_2 = 0.13 \times 10^{-2}$ for the silicone with shore hardness 28 A. Figure 6.12 reports the results of the FEM simulations. The 3D model is constituted by two elements made of silicone with shore hardness 28 A, by one layer of silicone with

Fig. 6.12 FEM analysis for the characterization of the second silicone molding
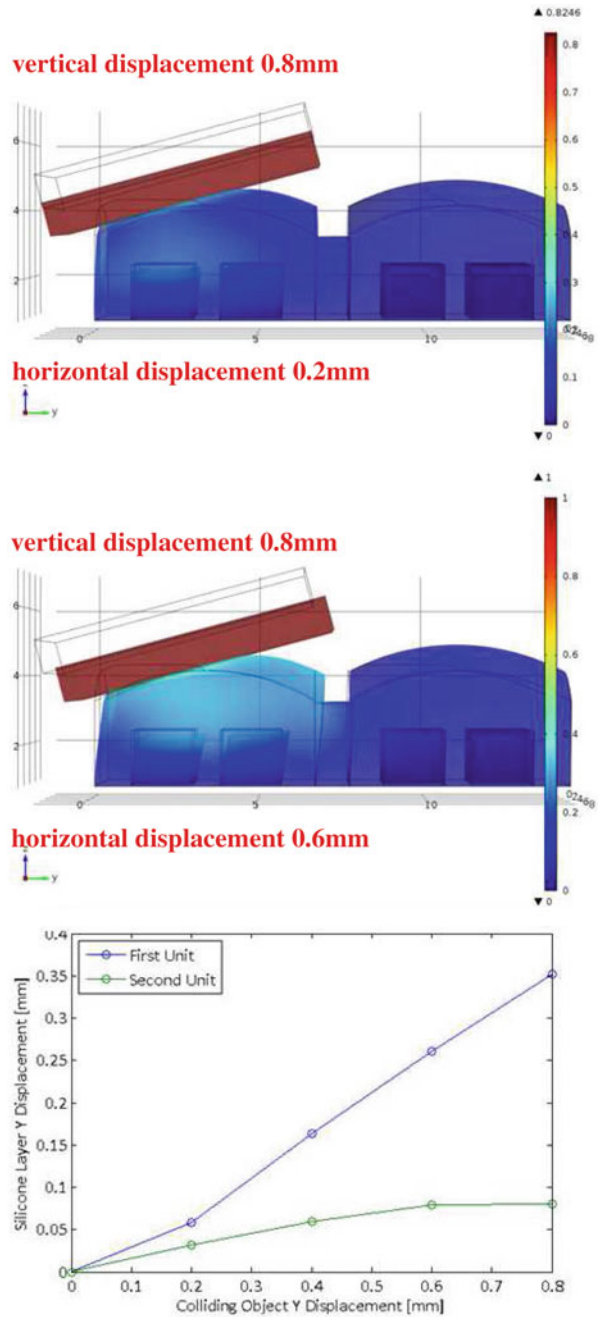
shore hardness 6 A positioned between the two modules and by a plane of aluminum material that represents the colliding object. The latter is subjected to a prescribed displacement chosen to generate a mechanical contact between the plane and one module resulting in a contact force. Referring to the reference frame reported in the figure, the displacement is applied only along the *y*- and *z*-axes. The applied vertical displacement is 0.8 mm and the horizontal displacement ranges from 0 to 0.6 mm with a step size of 0.2 mm. The mesh geometry is uniform for all the 3D model parts and it consists of 9469 tetrahedral elements. The last picture of Fig. 6.12 reports a comparison between the displacements along the y-axis that affects the two sensor deformable layers. The results show that for a 0.35 mm deformation of the first sensing module the corresponding maximum deformation on the second module is only 0.08 mm, that means a negligible mechanical coupling.

4.3 Integration of the Conformable Sensor on a Robot Arm

The first step towards the integration of the sensor patch on a robot link consists in bonding the deformable layer to the flexible PCB. In order to ensure correct operation of the sensor, the flexible PCB should be conformed to the surface selected for the final assembly of the sensor patch before silicone layer bonding. Since the force estimation depends on the deformations of the silicone layer, if the flexible PCB were conformed to the target shape after bonding of the deformable layer, a residual strain would affect the sensing modules causing a wrong force estimation. Therefore, reduction of these undesired deformations is achieved by first curving the PCB and then by bonding the silicone caps to it. To this aim, a mechanical support, designed on the basis of the shape of the surface selected for the final mounting, has to be realized. To mount the sensor patch on a KUKA LWR 4+, the support has been designed based on a 3D CAD model of the robotic arm and it has been 3D printed in ABS. The 3D CAD model provides a simple mechanism to extract a part of the robot surface where to collocate the distributed sensor. Figure 6.13 shows how
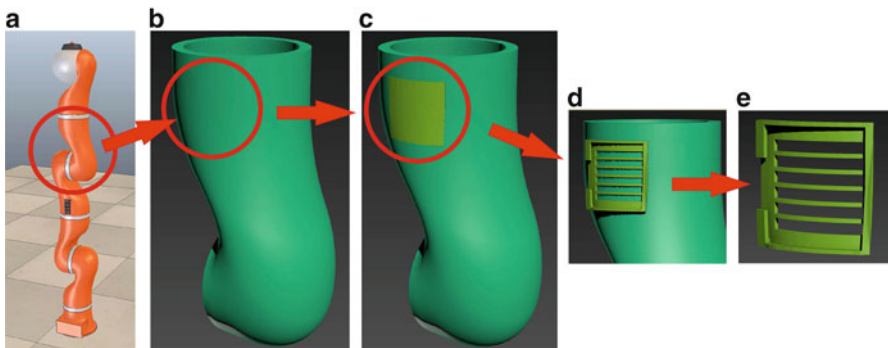

Fig. 6.13 Sensor patch mechanical support: (**a**) KUKA LWR 4+ 3D model, (**b**) selected surface, (**c**) extruded shape, (**d**) designed support on robot link, (**e**) designed support
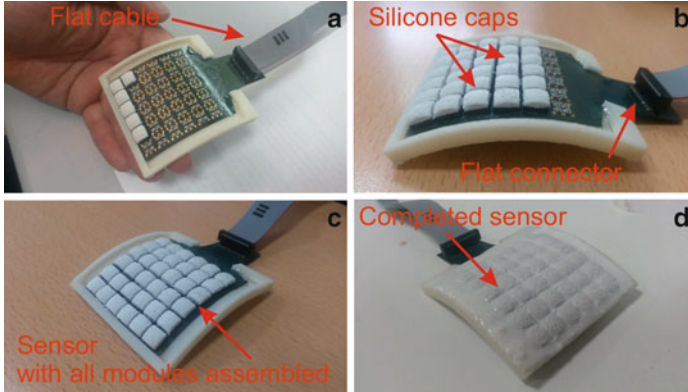
Fig. 6.14 Conformable sensor patch during assembly phases: (**a**) perspective view, (**b**) side view, (**c**) perspective view of the sensor prototype, (**d**) completed sensor prototype after second silicone molding.

the sensor patch support has been designed in the 3D CAD software. Starting from the CAD model of the whole arm (see Fig. 6.13a), the area identified for the final mounting of the sensor patch is selected (see Fig. 6.13b). From the selected surface, the sensor patch support has been extruded as a filled solid (see Fig. 6.13c). To obtain the final support, the filled solid has been completed with side edges designed to mechanically block the sensor patch on it and also several holes for inspection of the bottom side of the electronic layer (see Fig. 6.13d and e).Once the flexible PCB has been fixed to the mechanical support by epoxy resin, the silicone caps are bonded to each sensing modules on the optoelectronic layer and the soft silicone molding is applied, by obtaining a fully assembled patch. Figure 6.14 shows some pictures of the conformable sensor prototype during the assembly phases.

Note that the final conformed patch has not a uniform spatial resolution for the force detection, which depends on the local curvature. Let r_{flat} denote the sensing modules distance before bonding of the PCB on the conformed mechanical support, that is equal to the spatial resolution of the flat patch, namely 7.4 mm. Moreover, let R denote the local curvature radius of the mechanical support and h_c the height of the silicone caps bonded on the PCB. Then, the spatial resolution of the conformed sensor patch locally varies in the range $\pm h_c r_{\text{flat}}/R$ from the flat value r_{flat}. In particular, considering a curvature radius of 30 mm being $h_c = 4$ mm and $r_{\text{flat}} = 7.4$ mm, the spatial resolution of the conformed sensor patch locally varies from 7.4 to 8.4 mm.

4.4 Characterization as Distributed Force/Tactile Sensor

To provide the user with a contact force vector and contact position expressed in a single reference frame fixed to the sensor, the calibration procedure has to take into account the orientation of suitable reference frames attached to each module
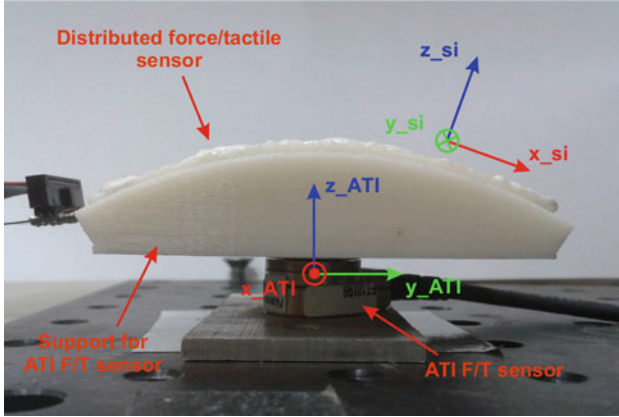
Fig. 6.15 Calibration setup

with respect to the given sensor frame. Furthermore, a comparison between two calibration algorithms with different computational complexity is reported: one based on an artificial neural network (ANN) and the other one based on a simple linear mapping.

The calibration procedure has been implemented by using the same reference sensor of Sect. 3.2. Figure 6.15 shows the calibration setup with the sensor patch mounted on a mechanical adapter fixed to the reference sensor. The computation of the calibration parameters for each sensing module requires that the axes of the reference frame of the ATI sensor and the axes of the reference frame of the sensing module are aligned. Let Σ_{s_i} be the reference frame of the ith sensing module and Σ_{ATI} the reference frame of the ATI sensor. Based on the 3D CAD models, the rotation matrix $\mathbf{R}_{s_i}^{ATI}$ for each sensing module can be defined. Then, with the choice of the reference frames reported in Fig. 6.15, the following equation, that allows to rotate the force vector measured by the ATI sensor in the ith sensing module reference frame, can be written as:

$$\mathbf{f}^{s_i} = \left(\mathbf{R}_{s_i}^{ATI}\right)^{\mathrm{T}} \mathbf{f}^{ATI}, \tag{6.10}$$

where $\mathbf{f}^{ATI} = (f_x^{ATI}\ f_y^{ATI}\ f_z^{ATI})^{\mathrm{T}}$ is the vector that contains the force components expressed in the ATI sensor frame, $\mathbf{f}^{s_i} = (f_x^{s_i}\ f_y^{s_i}\ f_z^{s_i})^{\mathrm{T}}$ is the vector that contains the force components expressed in the ith sensing module frame, and the rotation matrix $\mathbf{R}_{s_i}^{ATI}$ is defined as

$$\mathbf{R}_{s_i}^{ATI} = \mathbf{R}_z\left(\pi/2\right)\mathbf{R}_y\left(\theta_i\right), \tag{6.11}$$

where θ_i is the angle between the y_{ATI} e x_{s_i} axes, positively defined for counter-clockwise rotation about the y_{ATI} axis, that can be extracted from the 3D CAD model of the mechanical support.

To collect data for the calibration of the sensor patch, an operator carried out at least two experiments for each sensing module, by using a stiff flat object to apply different external forces. In particular, the operator manually interacted with each sensing module being careful to apply forces with components along all directions and with amplitudes up to the sensor full scale. For each experiment, all the voltage variations $\mathbf{V}^{s_i} = (V_1^{s_i}\ V_2^{s_i}\ V_3^{s_i}\ V_4^{s_i})^{\mathrm{T}}$ measured by the PTs belonging to the sensing module, and the force components $\mathbf{f}^{ATI}$ measured by the reference sensor have been acquired. Then, for each experiment the measured $\mathbf{f}^{ATI}$ have been rotated according to (6.10) in order to obtain the force components $\mathbf{f}^{s_i}$ in the sensor module frame. The collected data have been divided into two sets: a training set used to identify the calibration parameters and a validation set used to validate the accuracy of the calibration. With these data two phenomenological models have been calibrated and compared.

The first model considers the force components as a quasi-linear combination of the measured voltages as

$$f_x^{s_i} = (\mathbf{k}_x^{s_i})^{\mathrm{T}}\mathbf{V}^{s_i} \tag{6.12}$$

$$f_y^{s_i} = (\mathbf{k}_y^{s_i})^{\mathrm{T}}\mathbf{V}^{s_i} \tag{6.13}$$

$$f_z^{s_i} = (\mathbf{k}_z^{s_i})^{\mathrm{T}}\mathbf{g}(\mathbf{V}^{s_i}), \tag{6.14}$$

where the vector function $\mathbf{g}(\cdot)$ is the absolute value applied to each component of the vector $\mathbf{V}^{s_i}$ and the three 4×1 vectors $\mathbf{k}_x^{s_i}$, $\mathbf{k}_y^{s_i}$, and $\mathbf{k}_z^{s_i}$ represent the calibration parameters to identify. Starting from the training set data, these parameters have been identified with a simple least square algorithm by inverting Eqs. (6.12), (6.13), and (6.14), respectively, written for each point of the training data set. Then, the identified parameters have been used to evaluate the accuracy of the calibration phase, by computing the estimated force components for the validation data set, as

$$\hat{f}_x^{s_i} = (\mathbf{k}_x^{s_i})^{\mathrm{T}}\mathbf{V}^{s_i} \tag{6.15}$$

$$\hat{f}_y^{s_i} = (\mathbf{k}_y^{s_i})^{\mathrm{T}}\mathbf{V}^{s_i} \tag{6.16}$$

$$\hat{f}_z^{s_i} = (\mathbf{k}_z^{s_i})^{\mathrm{T}}\mathbf{g}(\mathbf{V}^{s_i}), \tag{6.17}$$

where $\hat{f}_x^{s_i}, \hat{f}_y^{s_i}$, and $\hat{f}_z^{s_i}$ are the estimated force values for the ith sensing module.

As second model, a $\mathbf{f}_{\mathrm{NN}}^{s_i}(\cdot)$ ANN operator has been considered, namely for the ith sensing module

$$\mathbf{f}^{s_i} = \mathbf{f}_{\mathrm{NN}}^{s_i}(\mathbf{V}^{s_i}). \tag{6.18}$$

A standard two-layer feed-forward neural network, trained with the Levenberg–Marquardt method, has been used to fit the training data. Different number of neurons of the hidden layer for the ANN have been tested, and the solution providing a good trade-off between the training error and complexity has been found with six neurons.

Fig. 6.16 Force components for a training data set: X component (*top*), Y component (*middle*), Z component (*bottom*)

The force components estimated by using both models have been compared to the measured ones. In Fig. 6.16, as an example, for a sensing module the estimated force components are compared to the actual force components measured by the reference sensor, for a training data set, just to verify the convergence of the training algorithm. Instead, Fig. 6.17 shows the accuracy of the calibration, by reporting, for the same sensing module, the estimated and the measured force components by using the validation data set. The estimation appears satisfactory for all force components with both calibration models.

In order to evaluate the calibration performance in a quantitative way, a synthetic index has been computed for each force component and for both models. The quality index is defined as

$$e_k = \frac{1}{N} \sum_{i=0}^{N} \left| f_k^{s_i}(i) - \hat{f}_k^{s_i}(i) \right|, \tag{6.19}$$

where $k = x, y, z$ indicates the force component, N is the number of samples, $f_k^{s_i}$ is the force component k of the ith sensing element, and $\hat{f}_k^{s_i}$ the corresponding
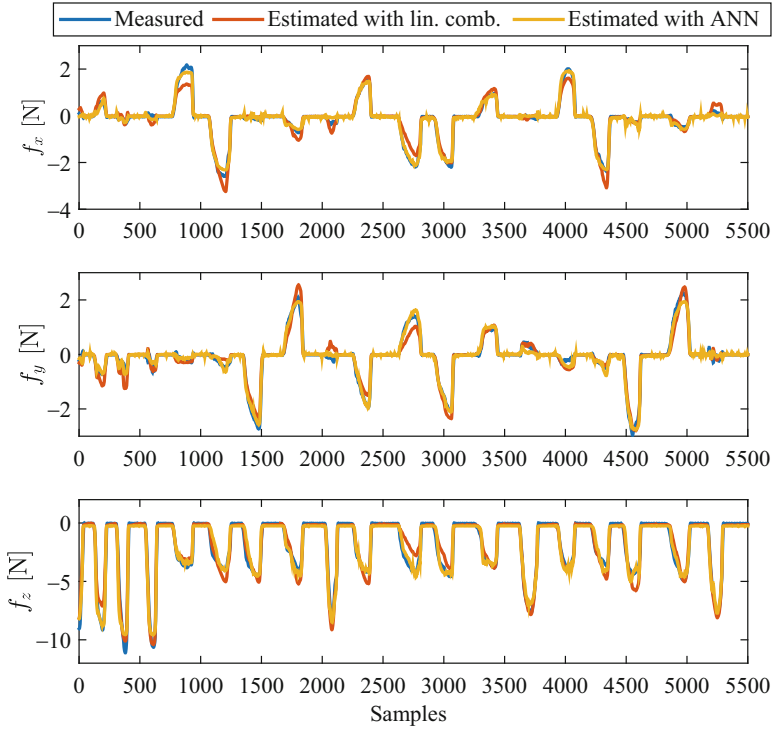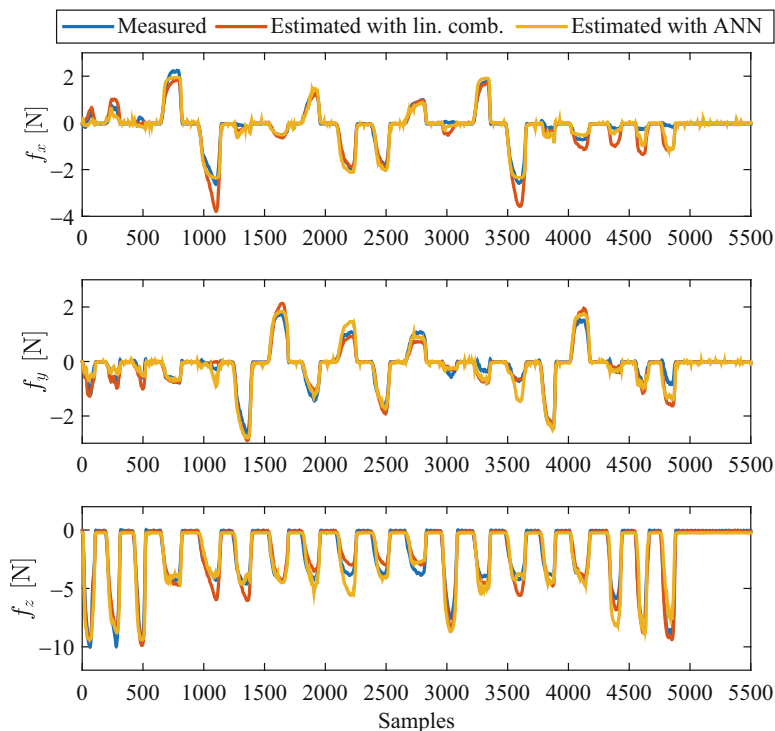
Fig. 6.17 Force components for a validation data set: X component (*top*), Y component (*middle*), Z component (*bottom*)

estimated value. The use of the calibration approach based on the ANN model provides a better accuracy for the shear force components for most of the sensing elements, but it introduces a greater mean error for the normal force component. Moreover, the better accuracy of the ANN model is not enough to justify its computational complexity compared to the quasi-linear model. By considering the full scales of each sensing element, which are $\pm 4\,\mathrm{N}$ for the shear components and $14\,\mathrm{N}$ for the normal component, the maximum mean error is less than $7.5\,\%$. Table 6.1 provides a comparison of the calibration approaches. It reports the mean estimation error and the standard deviation for each calibration approach computed by considering the 36 sensing modules. The absolute accuracy of the shear components is less than the one of the normal component, but taking into account the force range previously described, the relative accuracy results to be equivalent for all the force components. Up to now, all modules have been calibrated one by one. Instead, by using the same calibration parameters for all sensing modules, estimated on a module located in the middle of the patch, a degradation of the accuracy is expected. To quantify it, the linear model is adopted for all sensing modules and also the performances obtained in this case are reported in Table 6.1.

Table 6.1 Comparison of the three calibration approaches in terms of mean estimation error and standard deviation

	Mean error [N]			Standard deviation [N]		
Approach	f_x	f_y	f_z	f_x	f_y	f_z
Lin. comb.	0.1469	0.1657	0.4139	0.0363	0.0460	0.1052
ANN	0.1357	0.1445	0.4290	0.0485	0.0457	0.1497
Lin. comb. (one calib. matrix)	0.3964	0.2728	0.7388	0.1293	0.1011	0.2483

The maximum mean error is less than 15 %, which is the double of the former calibration approach, but using a significantly less time-consuming calibration procedure. It is evident that for whole body applications, a good trade-off can be: the use of the first strategy (more accurate and time-consuming) for parts (e.g., arms, hands) where the interactions with the environment are frequent and intentional; the use of the second approach (less accurate but time-saving) for parts (e.g., torso, legs) where typically only unintentional interactions can occasionally occur.

5 Experiments

This section reports experimental validations for both the six-axis force/tactile sensor and the distributed force/tactile sensor. The tactile sensor, mounted on a parallel gripper of a KUKA youBot, has been successfully exploited to detect and avoid slipping of objects in manipulation tasks. The distributed force/tactile sensor has been installed on a KUKA LWR4+ robot and the force estimation accuracy has been validated and compared with an indirect estimation technique.

5.1 Use of the Tactile Sensor for Slipping Detection and Avoidance

The slipping avoidance algorithm exploiting the six-axis force/tactile sensor comprises two control actions. The first action is based on the static Coulomb friction law and the second one is aimed at counteracting the uncertainties on the estimated friction coefficient and the object weight, as well as at counteracting time-varying external disturbances applied to the object.

Let f_n be the normal component of the force to grip the object without slipping when a tangential external force f_t is applied to it. The Coulomb friction law states that, in static conditions, slipping cannot happen if the normal force in (6.20) is applied to the object, where μ is the friction coefficient at the contact,

$$f_n = f_t/\mu. \tag{6.20}$$

With reference to Fig. 6.7, since the z axis of the sensor frame, aligned with the actuation direction of the gripper, forms an angle θ with the normal to the contact plane, to compute the correct actuation force u, the normal force has to be projected along the actuation direction through the angle θ, estimated on-line based on the tactile map, i.e.,

$$u = f_n/\cos\theta. \tag{6.21}$$

Thus, the static component of the control force is

$$u_s = f_t/(\mu\cos\theta), \tag{6.22}$$

where the tangential force can be computed by (6.7) on the basis of the force vector measured by the sensor.

Such control approach needs an estimated friction coefficient μ, which can be obtained through a preliminary exploration phase, described in the next section.

5.1.1 The Exploration Phase

The exploration procedure consists in applying a fixed normal force while slowly sliding the sensor over the surface of the object placed on a desk. The proposed procedure has been tested on different objects: an aluminum box with 80 g weight, a gum box with 35 g weight with and without the sealing plastic film (see Fig. 6.18).

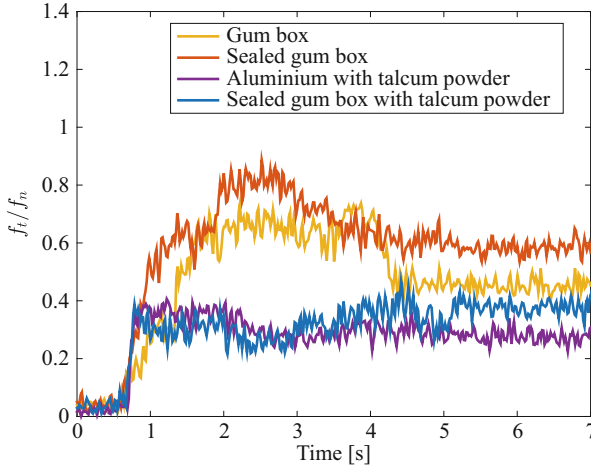**Fig. 6.18** The set of manipulated objects

Fig. 6.19 Tactile exploration of all objects in different surface conditions

The procedure has been repeated for all the objects, both when their surface is clean and for some of them when the surface is soiled by talcum powder.

Figure 6.19 reports the estimated friction coefficients. It clearly shows how the friction coefficient can significantly vary from material to material and, for the same material, depending on the conditions of its surface. In fact, the estimated friction coefficient varies from $\mu = 0.25$ of the aluminum box with talcum powder to $\mu = 0.6$ of the clean sealed gum box.

5.1.2   Slipping Avoidance in Dynamic Conditions

In the real-world, a slipping avoidance strategy based only on the estimated friction coefficient likely fails since the friction coefficient can change after the exploration phase. Moreover, there could be cases when the exploration cannot be performed.

To increase the robustness of the manipulation, an additional control strategy has been proposed by the authors in [18] and here briefly recalled, which exploits the six-axis force/tactile sensor.

According to the approach originally presented in [6], a linear Kalman filter (KF) is used to compute a control action, additional to the static one in (6.22), aimed at both counteracting uncertainties and tackling low breakaway forces due to time-varying external disturbance forces applied to the object.

The control law is an integral action on the KF residual, implemented using Tustin's method as reported in (6.24), where x_h is the integrator state at the hth time instant, k_c is the control gain, T is the sampling time, and f_{n_d} is the proposed slipping avoidance control action.

$$x_{h+1} = x_h + e_h \tag{6.23}$$

$$f_{n_d h} = k_c |Tx_h + T/2e_h| \tag{6.24}$$

The absolute value ensures that the normal force always pushes and the integrator ensures that $f_{n_d} = 0$ when e_h has zero mean.

In conclusion, the total actuation force can be computed as the sum of two contributions: the static control action in (6.22) and the slipping avoidance control action in (6.24), i.e.,

$$u_h = f_{th}/(\mu \cos \theta_h) + f_{n_d h}/\cos \theta_h \tag{6.25}$$

Note that the computation of the correct actuation force along the gripper motion direction requires estimation of the contact plane orientation angle θ_h, that can be performed using the tactile map, as explained in Sect. 3.2.

5.1.3 Results

The force/tactile sensor described in Sect. 3.2 has been integrated in the parallel gripper of a KUKA youBot (Fig. 6.20) by means of a USB interface and an ROS node publishing force/tactile data.

The slipping avoidance control algorithm has been implemented as another ROS node executed at a rate of 10 Hz. This low sampling rate is due only to the hw/sw architecture limitations of the setup, introduced by the installed version of the robot firmware and by the used ROS Fuerte wrapper that limits the general performance of the robot in tracking position and velocity commands. Nevertheless, satisfactory results have been achieved.
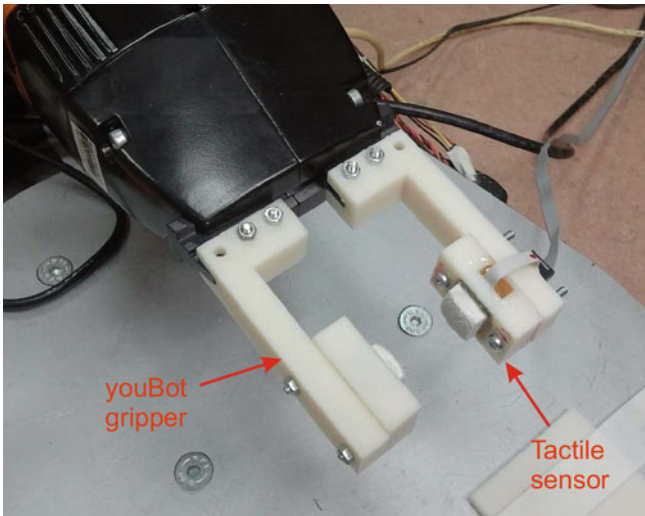

Fig. 6.20 Force/tactile sensor mounted on the parallel gripper of the youBot

The sensor data can be actually exploited to realize an effective slipping avoidance strategy an example of object manipulation in presence of uncertainties is here presented. Further experiments can be found in [18].

The experiment is aimed at lifting the aluminum box while proving the robustness of the slipping control. The estimated friction coefficient $\mu = 0.4$ is used in (6.25), but the surface of the aluminum box has been corrupted with talcum powder, which causes a significant drop of the actual friction coefficient (about 0.25 as seen in Fig. 6.19). Even though the weight has been considered known, the error on the estimation of the initial grip force is due to the error on the estimated friction coefficient that leads to an initial grasp force insufficient to hold the object with a dirty surface. Figure 6.21 reports the results of this experiment executed by deactivating the dynamic control action. The measured tangential component is lower than half of the object weight and it demonstrates that the robot is not able to lift the object, which slides between the gripper fingers resting on the plate. By activating the control action in (6.24) with $k_c = 8$, no slip occurs owing to the correction to the actuation force computed by the controller (yellow curve), as reported in Fig. 6.22.
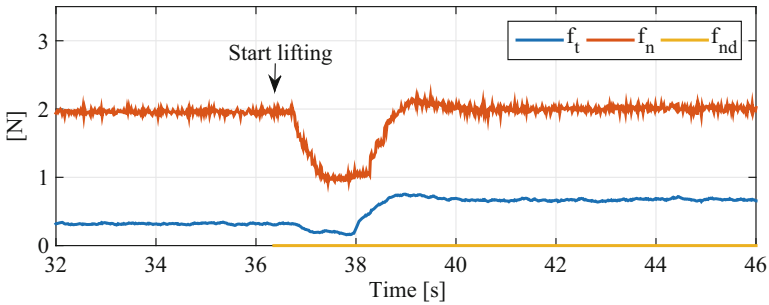

Fig. 6.21 Grasping of the aluminum box in perturbed conditions (changed friction coefficient) without dynamic slipping control
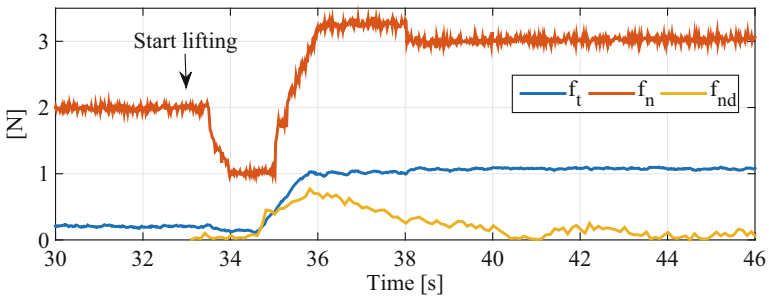

Fig. 6.22 Grasping of the aluminum box in perturbed conditions (changed friction coefficient) with dynamic slipping control

5.2   The Distributed Force/Tactile Sensor for Physical Human–Robot Interaction

In applications where the contact is intentional (advanced programming methods, interaction with the environment, and manipulation), contact points are typically located on the links and an accurate estimation of both the location and contact force vector is needed for a proper motion of the robot, e.g., allowing the user to move the robot links along specific Cartesian directions and improving the intuitiveness of the collaboration task. Instead, for handling unintentional collisions on other parts of the robot body a more rough information on contact location, but an accurate detection of the direction is needed for ensuring a safe behavior of the robot, i.e., quickly moving the arm away from the operator who touched the robot.

The aim of this section is to provide a validation of the calibration procedure described in Sect. 4.4 by testing the force/tactile sensor when mounted on a real robot. In particular, the more accurate, even though time consuming, calibration procedure that requires the calibration of each sensing module has been selected for comparison purposes. The robot selected for the analysis is a KUKA 7 DOF LWR4+ light-weight robot. The sensor is connected to an acquisition board with a flat cable long enough to avoid to install it on the robot and the sensor voltages are acquired through the "scanning strategy" detailed in Sect. 4.2. The board sends over a USB connection the acquired raw data to a PC running a C++ sensor library for interfacing with the data acquisition board. The sensor library is able to provide, on the basis of the sensor voltages, the estimated force vectors applied to the 36 sensor modules, the contact points and the contact frames. The information are sent via a UDP socket to a second PC used to execute the robot control algorithm.

5.2.1   Recall to the Residual-Based Technique for External Torques Estimation

In the experiments described below, two sensor patches are installed on the robot, one on the end effector together with an ATI F/T sensor and the other one on the third link (see Fig. 6.23). Several times, a force is applied by an operator to the force/tactile sensor on a single contact point while the robot is fixed in a given joint configuration, so as to allow a fair comparison with the contact force estimated using the residual-based method proposed in [21]. For the sake of completeness, such approach is briefly illustrated below. Consider the robot dynamic model, neglecting joint friction torques,

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}) = \boldsymbol{\tau} + \boldsymbol{\tau}_{\text{ext}}, \qquad (6.26)$$

where $\mathbf{q} \in \mathbb{R}^7$ is the vector of the generalized coordinates, $\mathbf{M}(\mathbf{q})$ is the symmetric, positive definite inertia matrix, $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})$ is the matrix of Coriolis and centrifugal terms, $\mathbf{g}(\mathbf{q})$ is the vector of gravity torques, $\boldsymbol{\tau}$ is the vector of control torques, and
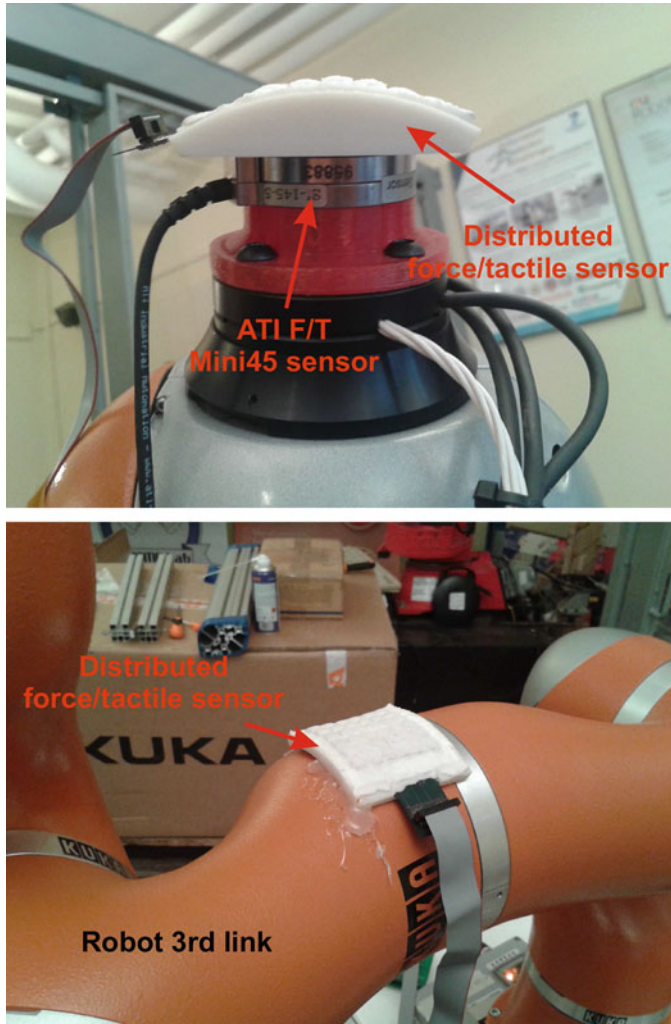
Fig. 6.23 The force/tactile sensor on the KUKA LWR 4+: the sensor on the end effector (*top*) and on the third link (*bottom*)

$\boldsymbol{\tau}_{\text{ext}}$ is the vector of torques due to external contact forces acting on the robot. The contact force estimation has been computed considering an approximate dynamic model of the KUKA LWR 4+ robot by neglecting friction torques.

Defining the generalized momentum of the robot as

$$\mathbf{p} = \mathbf{M}(\mathbf{q})\,\dot{\mathbf{q}},\tag{6.27}$$

the residual vector $\mathbf{r} \in \mathbb{R}^7$ can be expressed as

$$\mathbf{r} = \mathbf{K}_I \left(\mathbf{p} - \int_0^t \left(\boldsymbol{\tau} - \mathbf{C}^T (\mathbf{q}, \dot{\mathbf{q}}) \, \dot{\mathbf{q}} - \mathbf{g} (\mathbf{q}) + \mathbf{r} \right) \mathrm{d}s \right), \tag{6.28}$$

with $\mathbf{r}|_{t=0} = 0$ and $\mathbf{K}_I$ a diagonal positive definite matrix. From Eqs. (6.26) and (6.27), the dynamics of $\mathbf{r}$ can be expressed as

$$\dot{\mathbf{r}} = \mathbf{K}_I \left(\boldsymbol{\tau}_{\text{ext}} - \mathbf{r} \right). \tag{6.29}$$

By choosing a gain $\mathbf{K}_I$ large enough, assuming perfect knowledge of the dynamic model parameters, the asymptotic solution of (6.28) is

$$\mathbf{r} = \boldsymbol{\tau}_{\text{ext}}. \tag{6.30}$$

If $\mathbf{F}_k \in \mathbb{R}^3$ is an external force applied to a generic point of the robot and $\mathbf{J}_k (\mathbf{q})$ is the 3×7 Jacobian matrix associated with the contact point on the robot, the external torques and, thus, the residual computed as in Eq. (6.30), are related to $\mathbf{F}_k$ as

$$\boldsymbol{\tau}_{\text{ext}} = \mathbf{J}_k^{\text{T}} (\mathbf{q}) \, \mathbf{F}_k. \tag{6.31}$$

By solving Eq. (6.31), the component of the contact force vector not laying in the null space of $\mathbf{J}_k^{\text{T}}(\mathbf{q})$ can be computed as

$$\widehat{\mathbf{F}}_k = \left(\mathbf{J}_k^{\text{T}} (\mathbf{q}) \right)^{\#} \boldsymbol{\tau}_{\text{ext}}, \tag{6.32}$$

where $\left(\mathbf{J}_k^{\text{T}} (\mathbf{q}) \right)^{\#}$ is the generalized inverse of the Jacobian transpose.

5.2.2 Results

The first experiment involves the force/tactile sensor installed on an ATI F/T Mini45 sensor mounted on the robot end effector. The force measured with the ATI F/T sensor has been used as ground-truth. Figure 6.24 reports the components of the force measured with the two sensors and the force components estimated with the residual-based method, while Table 6.2 reports the mean errors computed by means of Eq. (6.19). With respect to the maximum value measured with the reference sensor for both the shear and normal components, the mean error for the tactile sensor is less than 5 %, while for the force estimated with the residual method it is about 30 %. The force estimated with the residual-based method is clearly less accurate than the force measured by the tactile sensor. The greater estimation error likely depends on the use of an uncertain and approximate dynamic model available in the KUKA FRI Library.

The second experiment involves the force/tactile sensor installed on the third link of the KUKA LWR 4+. In this case study, the force measured with the proposed
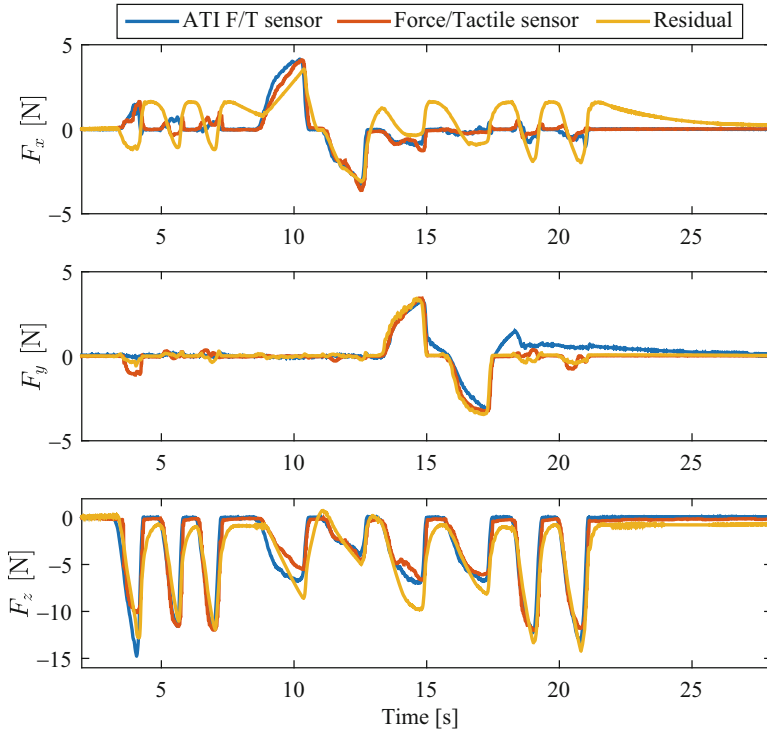
Fig. 6.24 Components of the force applied to the robot end effector expressed in base frame

Table 6.2 Estimation errors with respect to the ATI F/T Mini45 sensor expressed in Newtons

Sensor	f_x err	f_y err	f_z err
Tactile sensor	0.1412	0.1136	0.5721
Residual	1.1818	0.7100	4.9900

sensor has been compared only with the force estimated with the residual-based method since it was impossible to install the ATI sensor on the robot link. The results are reported in Fig. 6.25, where large errors on the residual-based estimation of the x component of the contact force are experienced. This is easily explained since the x axis of the base frame, in which the force vectors are expressed, lies along the direction parallel to the third link, which is in a configuration such that in this direction external forces are balanced by the mechanical structure of the arm and not by the joint torques. Also the y component is not perfectly estimated by the residual-based method even tough this component does not belong to the null space of the Jacobian transpose. From a safety point of view, this is a nice feature of the force/tactile sensor, since it allows the robot to detect contacts that cannot be detected by the other method. Moreover, since the sensor provides a direct measurement of the contact force, its installation on different parts of the robot structure does not affect the estimation accuracy. Differently, as reported in [39], the closer to the robot base is the contact point, the greater is the error of the residual
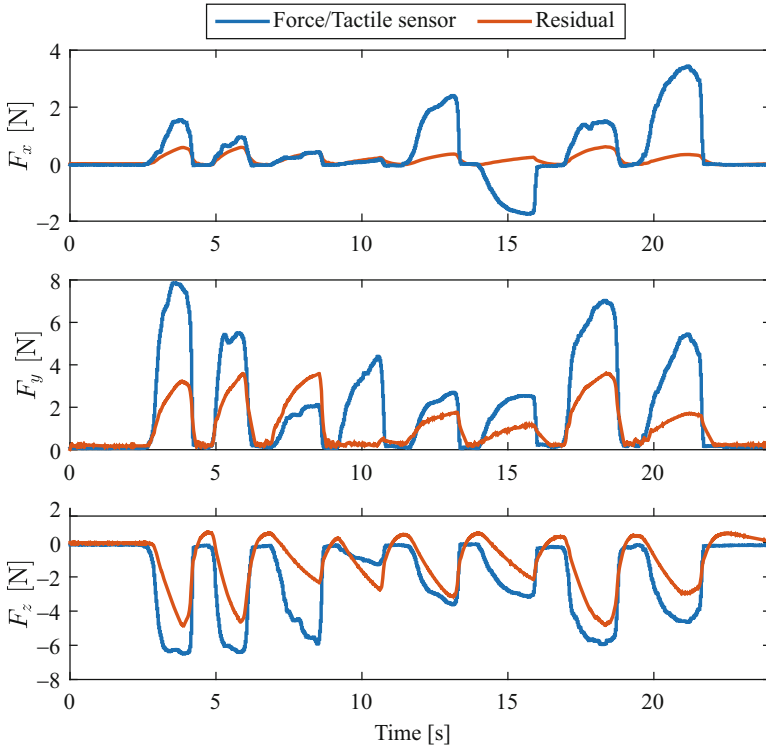
Fig. 6.25 Components of the force applied to the third link of the robot expressed in the base frame

estimation. Moreover, if three or more contacts occur on the structure of a robot with seven or less DOF, it is not possible to estimate the corresponding external forces, but just information on the external joint torques can be obtained. On the other hand, if a specific application requires a higher full scale range for the tactile sensor the mechanical structure of the latter, e.g., the deformable layer, would have to be redesigned by using a silicone with a higher shore hardness. This issue does not affect the residual-based method since the full scale of the estimated residual depends on the limits of the robot structure and sensors.

6 Conclusions

This chapter represents an overview of the authors' work on tactile sensing. Design and development of force/tactile sensors for fine manipulation, grasp, and safe human–robot interaction have been presented and experimentally tested. Firstly, a small tactile sensor able to measure the force and torque vectors that act on it has been presented. The sensor can be easily installed on a fingertip of a robotic hand

due to its shape and limited size. Its capabilities have been validated by executing a manipulation task where slipping detection and avoidance algorithms allowed to manipulate an object of unknown weight subject to dynamic disturbances. Secondly, design and implementation of a distributed, conformable force/tactile sensor have been reported. The authors have shown how the sensor is capable to estimate the force vector applied on the whole sensor surface with a uniform and fine accuracy demonstrating that the proposed solution represents an innovative and reliable way to provide the sense of touch to the whole robot body.

References

1. Alirezaei H, Nagakubo A, Kuniyoshi Y (2009) A tactile distribution sensor which enables stable measurement under high and dynamic stretch. In: IEEE symposium on 3D user interfaces, pp 87–93

2. Amarasinghe YWR, Kulasekera AL, Priyadarshana TGP (2013) Quantum tunneling composite (QTC) based tactile sensor array for dynamic pressure distribution measurement. In: In 2013 seventh international conference on sensing technology, pp 1–4

3. Ascari L, Corradi P, Beccai L, Laschi C (2007) A miniaturized and flexible optoelectronic sensing system for tactile skin. J Micromech Microeng 17:2288–2298

4. Beccai L, Roccella S, Arena A, Valvo F, Valdastri P, Menciassi A, Carrozza MC, Dario P (2005) Design and fabrication of a hybrid silicon three-axial force sensor for biomechanical applications. Sens Actuators A Phys 120:370–382

5. Bekhti R, Duchaine V, Cardou P (2014) Miniature capacitive three-axis force sensor. In: IEEE international conference on robots and systems, pp 3939–3946

6. Cavallo A, De Maria G, Natale C, Pirozzi S (2014) Slipping detection and avoidance based on Kalman filter. Mechatronics 24:489–499

7. Chi Z, Shida K (2004) A new multifunctional tactile sensor for three dimensional force measurement. Sens Actuators A Phys 111:172–179

8. Chu Z, Sarro PM, Middelhoek S (1995) Silicon three-axial tactile sensor. In: Solid-state sensors and actuators, pp 656–659

9. Cirillo A, Cirillo P, De Maria G, Natale C, Pirozzi S (2014) An artificial skin based on optoelectronic technology. Sens Actuators A Phys 212:110–122

10. Cirillo A, Cirillo P, De Maria G, Natale C, Pirozzi S (2014) A FE analysis of a silicone deformable interface for distributed force sensors. In: 7th international conference on times of polimers & composites, pp 485–488

11. Cirillo A, Ficuciello F, Natale C, Pirozzi S, Villani L (2015) A conformable force/tactile skin for physical human-robot interaction. IEEE Robot Autom Lett. doi:10.1109/LRA.2015.2505061

12. Crone B, Dodabalapur A, Lin YY, Filas RW, Bao Z, LaDuca A, Sarpeshkar R, Katz HE, Li W (2000) Large-scale complementary integrated circuits based on organic transistors. Nature 403:521–523

13. Cuellar F, Yamamoto T, Ishiguro H (2014) Design and development of a low power tactile multi-sensor network for robotic systems. In: International conference on mechatronics and automation 331–336

14. D'Amore A, De Maria G, Grassia L, Natale C, Pirozzi S (2011) Silicone-rubber-based tactile sensors for the measurement of normal and tangential components of the contact force. J Appl Polym Sci 122:3758–3770

15. Dahiya RS, Metta G, Valle M, Sandini G (2010) Tactile sensing - from humans to humanoids. IEEE Trans Robot 26:1–20

16. De Maria G, Natale N, Pirozzi S (2012) Force/tactile sensor for robotic applications. Sens Actuators A Phys 175:60–72
17. De Maria G, Natale C, Pirozzi S (2013) Tactile data modeling and interpretation for stable grasping and manipulation. Robot Auton Syst 61:1008–1020
18. De Maria G, Falco P, Natale C, Pirozzi S (2015) Integrated force/tactile sensing: the enabling technology for slipping detection and avoidance. In: IEEE international conference on robotics and automation, pp 3883–3889
19. Duchaine V, Lauzier N, Baril M, Lacasse MA, Gosselin C (2009) A flexible robot skin for safe physical human robot interaction. In: IEEE international conference on robotics and automation, pp 3676–3681
20. Elkmann N, Fritzsche M, Schulenbur E (2011) Tactile sensing for safe physical human-robot interaction. In: International conference on advances in computer-human interactions, pp 212–217
21. Flacco F, De Luca A (2012) Integrated control for pHRI: collision avoidance, detection, reaction and collaboration. In: Fourth IEEE RAS/EMBS international conference on biomedical robotics and biomechatronics, pp 288–295
22. Gerratt AP, Sommer N, Lacour SP, Billard A (2014) Stretchable capacitive tactile skin on humanoid robot fingers - first experiments and results. In: 14th IEEE-RAS international conference on humanoid robots, pp 238–245
23. Golger D, Gorges N, Worn H (2009) Tactile sensing for an anthropomorphic robotic hand: hardware and signal processing. In: IEEE international conference on robotics and automation, pp 895–901
24. Guizzo E, Ackerman E (2012) The rise of the robot worker. IEEE Spectr 49:34–41
25. Hammock ML, Chortos A, Tee BCK, Tok JBH, Bao Z (2013) 25th anniversary article: the evolution of electronic skin (eskin): a brief history, design considerations, and recent progress. Adv Mater 25:5997–6038
26. Hellard G, Russell RA (2002) A robust, sensitive and economical tactile sensor for a robotic manipulator. In: Proceedings of Australasian conference on robotics and automation, pp 100–104
27. Heo JS, Chung JH, Lee JJ (2006) Tactile sensor arrays using fiber Bragg grating sensors. Sens Actuators A Phys 126:312–327
28. Horii T, Nagai Y, Natale L, Giovannini F, Metta G, Asada M (2014) Compensation for tactile hysteresis using gaussian process with sensory Markov property. In: 14th IEEE/RAS international conference on humanoid robots, pp 993–998
29. Hoshi T, Shinoda H (2006) Robot skin based on touch-area-sensitive tactile element. In: IEEE international conference on robotics and automation, pp 3463–3468
30. Hwang ES, Seo JH, Kim YJ (2007) A polymer-based flexible tactile sensor for both normal and shear load detections and its application for robotics. IEEE/ASME J Microelectromech Syst 16:556–563
31. Iwasaki T, Takeshita T, Arinaga Y, Uemura K, Ando H, Takeuchi S, Furue M, Higurashi E, Sawada R (2015) Shearing force measurement device with a built-in integrated micro displacement sensor. Sens Actuators A Phys 221:1–8
32. Jiang F, Tai YC, Walsh K, Tsao T, Lee GB, Ho CM (1997) A flexible MEMS technology and its first application to shear stress sensor skin. In: IEEE Proceedings of 10th annual international workshop on micro electro mechanical systems, pp 465–470
33. Kamiyama K, Vlack K, Mizota T, Kajimoto H, Kawakami N, Tachi S (2005) Vision-based sensor for real-time measuring of surface traction fields. IEEE Comput Graph Appl 25:68–75
34. Kasap SO (2001) Optoelectronics and photonics: principles and practices. 1st edn. Prentice Hall, Upper Saddle River
35. Krishna GM, Rajanna K (2004) Tactile sensor based on piezoelectric resonance. IEEE Sens J 4:691–697
36. Lamy X, Colledani F, Geffard F, Measson Y, Morel G (2009) Robot skin structure and performances for industrial robot comanipulation. In: IEEE international conference on advanced intelligent mechatronics, pp 427–432

37. Lee MH (2000) Tactile sensing: new directions, new challenges. Int J Robot Res 19:636–643
38. Liu T, Inoue Y, Shibata K (2009) A small and low-cost 3-d tactile sensor for a wearable force plate. IEEE Sens J 9:1103–1110
39. Magrini E, Flacco F, De Luca A (2014) Estimation of contact forces using a virtual force sensor. In: IEEE/RSJ international conference on intelligent robots and systems, pp 2126–2133
40. Melchiorri C, Moriello L, Palli G, Scarcia U (2014) A new force/torque sensor for robotic applications based on optoelectronic components. In: IEEE international conference on robotics and automation, pp 6408–6413
41. Missinne J, Bosman E, Van Hoe B, Van Steenberge G, Kalathimekkad S, Van Daele P, Vanfleteren J (2011) Flexible shear sensor based on embedded optoelectronic components. IEEE Photon Technol Lett 23:771–773
42. Motoo K, Arai F, Fukuda T (2007) Piezoelectric vibration-type tactile sensor using elasticity and viscosity change of structure. IEEE Sens J 7:1044–1051
43. Mukai T, Onishi M, Odashima T, Hirano S, Luo Z (2008) Development of the tactile sensor system of a human-interactive robot (RI-MAN). IEEE Trans Robot 24:505–512
44. Narendiran A, George B (2015) Capacitive tactile sensor with slip detection capabilities for robotic applications. In: IEEE international instrumentation and measurement technology conference, pp 464–469
45. Novak JL (1989) Initial design and analysis of a capacitive sensor for shear and normal force measurement. In: Proceedings of IEEE international conference on robotic and automation, pp 137–145
46. Ohka M, Mitsuya Y, Matsunaga Y, Takeuchi S (2004) Sensing characteristics of an optical three-axis tactile sensor under combined loading. Robotica 22:213–221
47. Ohmura Y, Kuniyoshi Y, Nagakubo A (2006) Conformable and scalable tactile sensor skin for curved surfaces. In: IEEE international conference on robotics and automation, pp 1348–1353
48. Petrina AM (2011) Advances in robotics. Autom Doc Math Ling 45:43–57
49. Qi H, Joyce K, Boyce MC (2003) Durometer hardness and the stress-strain behavior of elastomeric materials. Rubber Chem Technol 76:419–435
50. Salo T, Vancura T, Baltes H (2006) CMOS-sealed membrane capacitors for medical tactile sensors. J Micromech Microeng 16:769–778
51. Schmitz A, Maggiali M, Natale L, Metta G (2010) Touch sensors for humanoid hands. In: IEEE RO-MAN 2010, pp 691–697
52. Schmitz A, Maggiali M, Natale L, Bonino B, Metta G (2010) A tactile sensor for the fingertips of the humanoid robot iCub. In: IEEE/RSJ international conference on intelligent robots and systems, pp 2212–2217
53. Shimojo M, Namiki A, Ishikawa M, Makino R, Mabuchi K (2004) A tactile sensor sheet using pressure conductive rubber with electrical-wires stitched method. IEEE Sens J 4:589–596
54. Siciliano B, Khatib O (2008) Springer handbook of robotics, 1st edn. Springer, Berlin
55. Strohmayr MW, Worn H, Hirzinger G (2013) The DLR artificial skin step I: Uniting sensitivity and collision tolerance. In: IEEE international conference on robotics and automation, pp 1004–1010
56. Tarchanidis KN, Lygouras JN (2003) Data glove with a force sensor. IEEE Trans Instrum Meas 52:9–989
57. Ulmen J, Cutkosky M (2010) A robust, low-cost and low-noise artificial skin for human-friendly robots. In: IEEE international conference on robotics and automation, pp 4836–4841
58. Wisitsoraat A, Patthanasetakul V, Lomas T, Tuantranont A (2007) Low cost thin film based piezoresistive MEMS tactile sensor. Sens Actuators A Phys 139:17–22
59. Yamada Y, Cutkosky MR (1994) Tactile sensor with 3-axis force and vibration sensing functions and its application to detect rotational slip. In: Proceedings of IEEE international conference on robotic and automation, pp 3550–3557
60. Zhang T, Liu H, Jiang L, Fan S, Yang J (2013) Development of a flexible 3-d tactile sensor system for anthropomorphic artificial hand. IEEE Sens J 13:510–518
61. Zhang T, Jiang L, Wu X, Feng W, Zhou D, Liu H (2015) Fingertip three-axis tactile sensor for multifingered grasping. IEEE/ASME Trans Mechatron 20:1875–1885

Chapter 7
Mechanical Characterization of MEMS

Seyed M. Allameh

Abstract Characterization of microelectromechanical systems, MEMS structures is important for their reliability, specifically for life-sustaining applications. Due to size effect, testing must be done at microscale under the same conditions the components are utilized. Microtesting systems have been developed and used for metallic and nonmetallic MEMS components. Other methods including on-the-chip characterization have been developed and utilized. Details of development or utilization of various micro-characterization techniques are presented in this chapter. There are many issues related to mechanical properties, some of which are briefly described here. Modeling MEMS both computationally and mathematically are depicted along with examples. At the end, future trends in MEMS are discussed.

Keywords MEMS • Characterization • Microtesting • Microtensile • Microfatigue • Failure • Microcracks • Reliability

1   Introduction

Microelectromechanical systems, MEMS devices are becoming ubiquitous, from automotive tire pressure sensors, inertial measurement units, and fuel injection nozzles to aerospace exploration devices. Other applications include tablets, mobile phones and watches, sport wearable gadgets, CO_2 and other gas sensors, and even blood viscosity measuring devices. In the light of these facts, special attention must be paid to MEMS characterization. Since mechanical properties are affected by a component's size, it is important to conduct characterization experiments at a comparable scale. The reliability of MEMS structures becomes vital for applications where human life is at stake. Examples are triggers for airbags and weaponry, life-supporting medical devices, and navigating systems for airplanes and vehicles. While macroscale mechanical characterization is well established, the processes at microscale yet have to be developed, enhanced, and standardized. Since MEMS

S.M. Allameh (✉)
Northern Kentucky University, Highland Heights, KY, 41099 USA
e-mail: allamehs1@nku.edu

field is relatively new, researchers are still trying to find more effective ways to characterize and model the behavior of MEMS structures.

To design, fabricate and successfully test MEMS components, an integrated approach is needed that provides Analysis, Computational and Experimental Solution(s) (ACES) [1]. This methodology employs analytical tools that are based on exact closed form solutions. Commonly used computation methods such as finite element method, FEM, boundary element method, BEM, and finite difference method, FDM, treat problems numerically, and provide approximate solutions. The degree of approximation, however, depends on the characteristics of the differential equations and the specifications of the boundary, initial and loading (BIL) conditions [1]. Unlike analytical or computational methods, experimental tools examine and evaluate actual objects, their performance, and their respective operating conditions and offer characteristics for those objects. The successful design and development of MEMS, with added complexity due to size effect, requires a combination of both experimental and theoretical tools [1].

2   Mechanical Properties-Related Issues

The movement of the mechanical components of MEMS is often associated with loads that can cause fatigue and failure of the systems. Mechanical properties such as strength, toughness, and fatigue life become important for the safe operation of MEMS. Devices such as micromirrors potentially have millions of mechanical components that tilt at high frequencies. Stresses developed in the hinges can lead to premature failure of one of these, possibly rendering the whole device useless. For bioMEMS that deal with various fluids, aggressive environments can lead to stress corrosion cracking. High-aspect-ratio surface-machined BioMEMS components such as microneedles and thermal-post heat-exchangers require good fracture toughness in addition to reasonable strength. Microengine components rely on good tribological properties. Adhesion properties of MEMS are important for microswitches. Thermal actuators operate at higher temperatures and therefore need good thermal-fatigue as well as creep resistance.

The effect of size on mechanical properties is well known for small structures. While elasticity modulus is not size-dependent (for same crystal orientations), other properties such as strength, fatigue limits, fatigue life, and hardness are size dependent. The effect becomes more pronounced when the size of the structure approaches the order of the grain size. Grain growth is one of the factors that affects the strength of the MEMS structures. Consider LIGA MEMS components being grown in PMMA molds. Initially, grains are very fine. This means a thin film, with a microstructure mostly made of these fine grains, will be pretty strong. This is not the case for thicker films where the fine grains were allowed to grow into coarse columnar grains, subsequently reducing the overall strength of the film. Such reduction in strength as well as in fatigue limit is well documented [2]. This is the case for most thin film depositions, lithography-based electrodepositions, and other additive processes.

Environmental effects on silicon greatly influences the reliability of MEMS devices. Stiction, specifically for MEMS moving components is one of the major failure causes. This happens to sliding surfaces that stick to each other during motion [3]. Presence of water vapor, not only promotes stiction, but also causes, fatigue damage accumulation [4] as well as stress corrosion cracking [5–7]. Stress-assisted dissolution of silica backbone at loci of high stresses (e.g., at the notch root) is postulated to be the mechanism for deepening the surface perturbations, leading to the formation of the grooves [4, 8–10]. Another mechanism for crack initiation has been suggested to be the thickening of the native silicon dioxide layer at locations where stress levels are high. Cracks have been suggested to start from these thick silicon dioxide regions [5].

Failure analysis of MEMS structures has been performed by optical and electron microscopy, atomic force microscopy, acoustic microscopy, focused ion beam microscopy, and scanning laser microscopy. It has been hard to duplicate failures in MEMS. This has been a major challenge in the failure analysis of these systems [11]. Reliability studies on silicon MEMS include the electrical actuation of comb drives. However, there are other characterization tools, such as in situ visual inspection of moving components, acquiring test data and tools for obtaining performance characteristics [12]. A number of reliability factors have been studied for silicon. These are frequency [13], shock resistance [14, 15], stiction [16], mechanical wear [17], fatigue [18, 19], fracture [20], vibration [21], and environmental effects [22]. MEMS devices that have been studied include microactuators [12, 23], microengines [13, 24], microrelays [25], thin films [26, 27], and resonating structures [6, 7] among others.

MEMS mechanical reliability studies have been conducted for decades [28, 29]. However, there are still many MEMS applications that are not possible because of the lack of mechanical behavior characteristics of such systems [30]. Although analytical models have been developed [31–35] to predict the reliability of MEMS or to help with reliability design rules, nevertheless, mechanical tests are still needed to provide data that validate such predictions.

Actuation of MEMS is achieved by different methods. Conventional approaches are simple and easy to implement and integrate into the circuits. There are still challenges associated with each of these methods. Electrostatically driven actuators are one-dimensional and effective. Examples are comb drives with interdigitating combs that can be used for both actuation and sensing. The main challenge is the high voltages (tens if not hundreds of volts) required for their actuation [36]. On the other hand, magnetically driven actuators use smaller voltages but higher currents that make them ineffective as thin films [37]. Despite their high energy densities, it is difficult to integrate them with conventional semiconductor processing techniques. When it comes to efficient actuators such as piezoelectrics, the challenge lies in their design for in-plane movement [38]. While yielding high forces, thermal actuators generate small displacements. Nevertheless for large displacements, they yield small forces making these actuators ineffective [39]. Efforts have been made to mitigate this problem by fabrication of a thermal actuator with medium range forces associated with medium range displacement [39]. In this design, two thermal

actuators generate rotation deflections that cause the motion of a cantilever at a voltage of 13.3 V and a current of 39 mA. The cantilever has a spring constant of 65.5 N/m and can deliver noticeable forces. Deflections of up to 28 μm and force efficiencies of up to 7 mW/μN have been reported [39].

3   Microtesting Systems

Microtensile Testing: The pioneering work of Sharpe and McAleavey [40] in developing microtensile and microfatigue testing systems was continued by other researchers who complemented these systems with new measurement techniques. Figure 7.1 presents the schematic of an early system that utilizes laser interferometry in measuring displacement.

The microtensile testing system utilizes various types of microsamples. One such microspecimen is a small dog-bone shape structure made of LIGA Ni MEMS deposited in PMMA molds. This structure has triangular ends, and its gage section measures 200 μm $\times$ 200 μm $\times$ 400–1000 μm. Both sides of the samples are polished to a mirror finish, and then marked in two locations by a pyramid indenter. These inverted pyramid indents are necessary for reflecting a laser beam leading to interferometry patters. Samples are mounted on small holders that are attached on one side to a fixed mount and on the other to the piston of an air bearing. The latter is part of a load train that connects the air bearing to a load cell and to a Uni-slide drive mechanism as shown in Fig. 7.1. The laser beam shines on the two inverted pyramids and generates reflected beams that interact and form interferometry patterns. These patterns are projected onto two photodiode arrays (detectors) mounted on two posts at the sides of the laser beam source. Light and dark bands of the pattern are analyzed
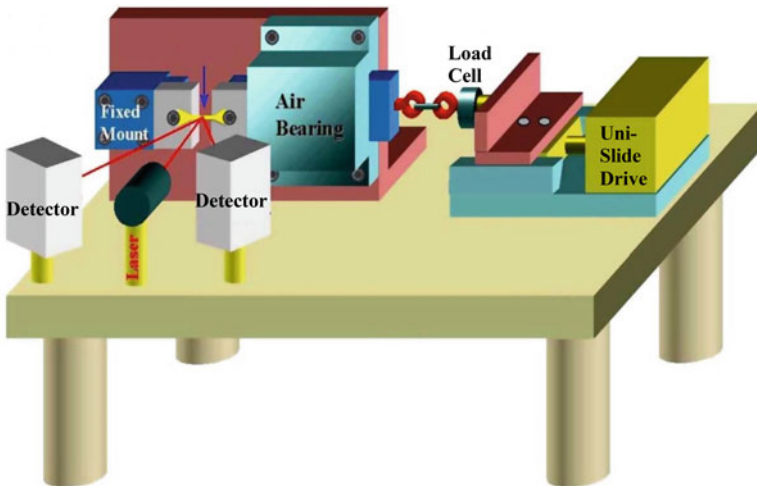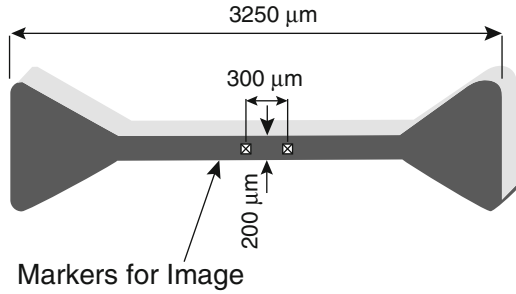

Fig. 7.1   Schematic of the microtensile testing system developed after Sharpe et al. [40]

Fig. 7.2 A typical dog-bone sample used in microtesting

and the change in fringes is recorded and used in the calculation of strain during tensile testing of the sample. The load cell records the load as the Uni-slide drive imposes strain on the microsample. A schematic of a dog-bone shape sample typical of tensile specimens is shown in Fig. 7.2.

Microfatigue Testing Systems: The microfatigue systems built based on Sharpe's initial design contain a piezo actuator capable of generating 0–180 μm of travel with frequencies up to 1000 Hz. The voltage applied to the piezo crystal is proportional to the displacement produced, leading to nanoscale motion. A small button-like load sensor is attached to the free end of the piezo drive, and a gripper is attached to the other side of the load cell. A second gripper is attached to a fixed mount. These grippers have recessed triangular shape areas that house the dog-bone shape end of the microsample. The displacement can be measured in two ways. One way is by recoding the data collected by a displacement monitor built into the piezo drive. The second method is by taking a series of pictures (e.g., a video clip) and analyzing them postmortem. Tests are conducted under load-control using LabVIEW™ (National Instruments Corp., Austin TX) or similar programs. A schematic picture of microfatigue system is presented in Fig. 7.3.

By changing the grippers, small-size CT samples can be tested in fatigue [41]. A typical small CT sample measuring 10 mm × 12 mm × 0.24 mm is shown in Fig. 7.4. The sample holder is in the form of a jacket that completely envelopes the sample except for the pin holes. It also has a small window on one side in its midsection, where the crack is expected to grow. This allows monitoring of the crack tip as it grows during fatigue testing. The jacket is necessary to prevent possible buckling of the thin CT sample.

More recently, a hybrid microtesting system has been developed to combine the functionalities of a microfatigue testing system with that of a microtensile one [42]. The piezo actuator used for nanoscale displacement amplitude of fatigue loading is mounted on a millimeter-scale Uni-slide drive. The latter movement is required for microtensile testing that allows small-size CT specimen characterization. Crack tip travel of this specimen is in the millimeter range (2–4 mm); however, fatigue loading will require displacement amplitudes that are in hundreds of nanometers. LabVIEW programming combined with special functions built into the piezo controller allow uninterrupted testing of the CT specimen (Fig. 7.4).

Fig. 7.3 Schematic of
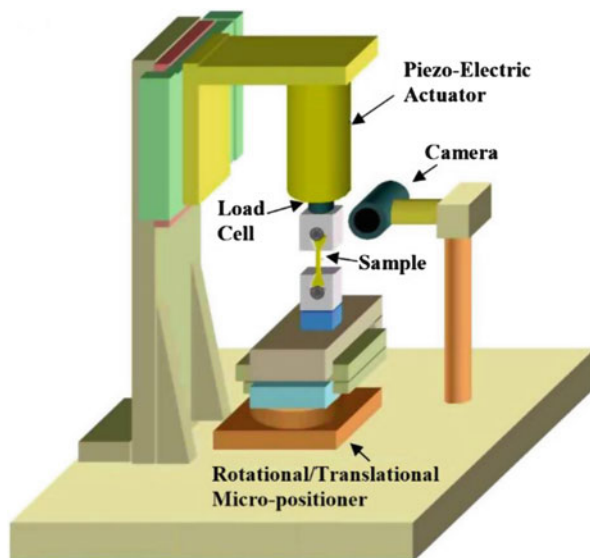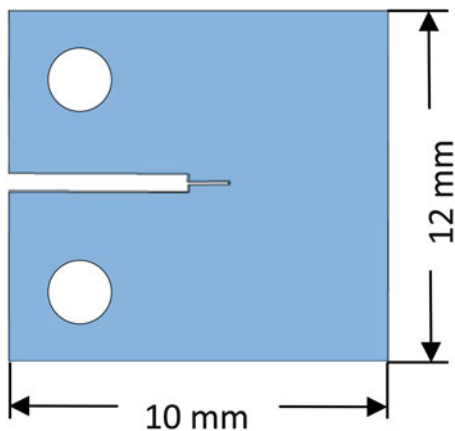microfatigue testing system

Fig. 7.4 Schematic of a
small-size thin CT LIGA Ni
sample

Microtesting systems have been used to investigate the properties of aluminum foam struts as well as natural fibers [43, 44]. Special fork-like grippers allow mounting Al struts that are cut from open cell aluminum foam block using a ram-type EDM. Strut properties can be incorporated into a mechanics model that predicts the mechanical properties of foam blocks [43, 45]. Local deformation of the struts is measured by image analysis of a series of pictures taken during tensile testing. These show the change in the distance between two landmarks created on the gage section before testing. Natural landmarks on the surface of the struts may also be used to determine the local deformation. Figure 7.6 shows how these landmarks are recognized for determination of their distances throughout the test.
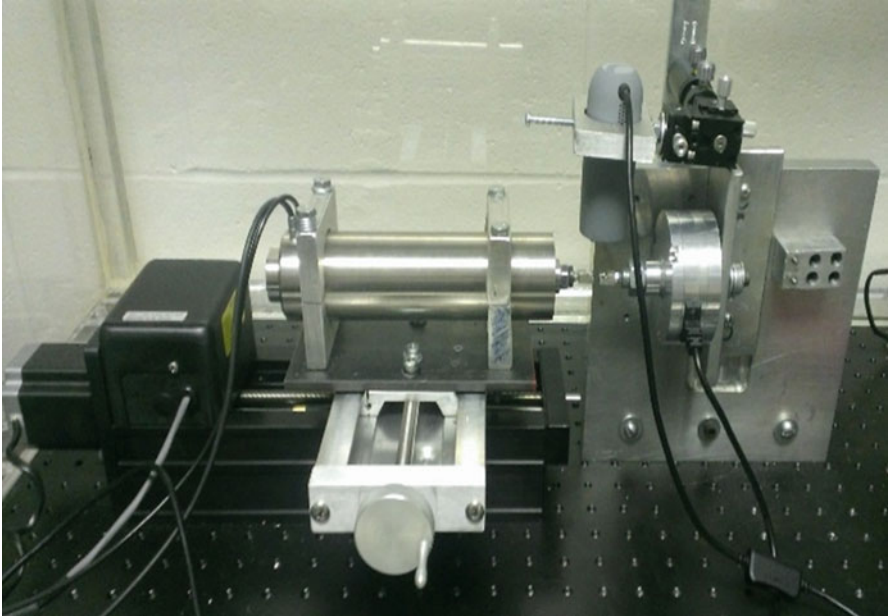
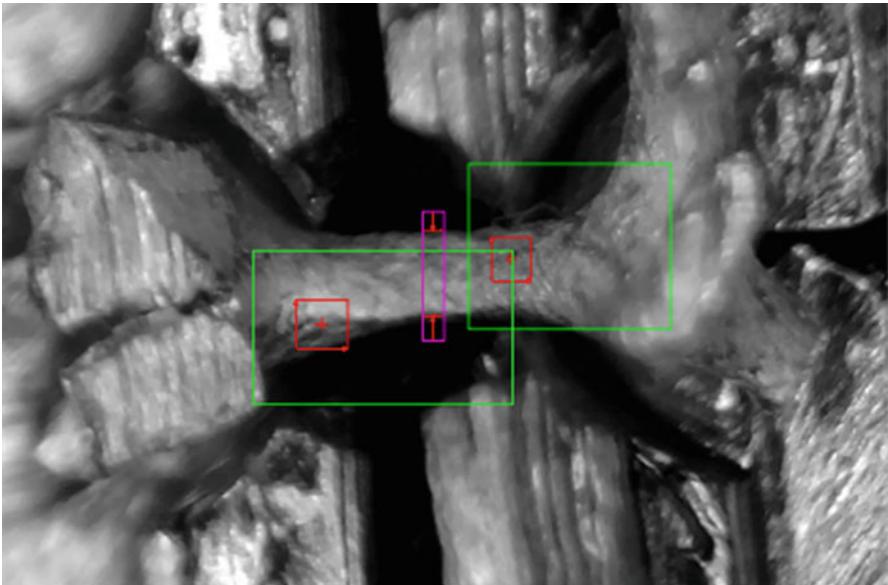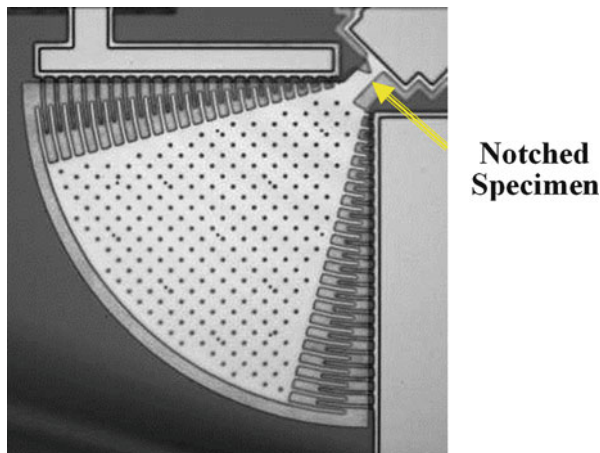**Fig. 7.5** A hybrid microtesting system combining micro- and nanoscale movement

Fig. 7.6 Aluminum strut picture being analyzed for local deformation

Fig. 7.7 An SEM image of
a silicon fatigue specimen

For testing brittle materials in fatigue, a special microfatigue system has been developed [46]. It consists of a capacitively driven comb drive attached to the fatigue specimen. Such a comb drive is shown (see Fig. 7.7) for silicon, measuring about $200\,\mu m \times 200\,\mu m$. Two interdigitating combs, one attached to the specimen and the other to the substrate, are energized by a sinusoidal voltage wave form at frequencies scaling with the natural frequency of the specimen. For the comb drive shown in Fig. 7.4, a function generator produces a 0–5 V signal that is augmented by an amplifier to 70–140 V and then applied to the interdigitating combs. The specimen with the comb structure constitutes a cantilever with the specimen being at the hinge. The oscillating cantilever will generate enough stresses at the hinge that cause it undergo fatigue and eventually fracture.

The response of the silicon cantilever to the applied voltage is not linear, so there is a need for a calibration curve that allows determination of deformation at various applied voltages. This is achieved by the use of image analyses of pictures obtained at different actuation voltages [46]. The analysis is based on block-matching and ingredient-based method [47–51]. To measure the level of stress at the root of the notch on the specimen, finite element analysis is performed. The elasticity modulus of the structures is usually known by various methods including microtensile testing. Using this known property and knowing the extent of angular displacement (by image analysis), the local stresses can be calculated for different loci on the sample. As an example, for an angular displacement of 1.44°, the local stress changes from 2.74 GPa tensile at the root of the notch to 122 MPa compressive at the back side of the sample (See the location in Fig. 7.7 where the arrow points to) [50].

When this test is conducted under the scanning tip of an AFM raster scanning the surface of the sample, the topography evolution of the surface can be studied. Using this method, the mechanisms of fatigue and fracture of silicon is investigated. Applying FFT transforms of the asperities associated with the roughness near the notch root, certain models can be considered to explain crack initiation from cusps that develop during fatigue of silicon [50].
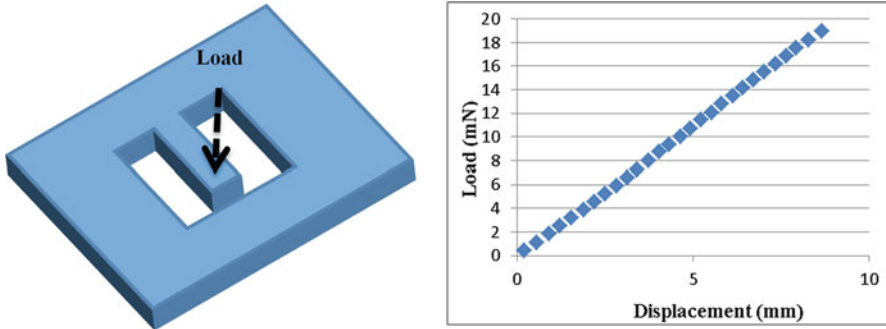
Fig. 7.8   Fracture toughness obtained from indentation of thin polysilicon film

Direct Loading Methods: Characterization of microsamples has been conducted using small loads, directly applied on specimens. These are usually cantilevers machined out of larger samples. An example is presented in Fig. 7.8, depicting a cantilever beam machined out of Ni-P plate with a thickness of 10 µm. The beam is notched to a depth of 3 µm at a location about 10 µm from the fixed end of the cantilever. The effective length of the beam is 40 µm. The results shown in Fig. 7.8 indicate linear elastic behavior until fracture (e.g., elastic-plastic deformation occurs only at the tip of the notch followed by an abrupt fracture). A Griffith fracture theory analysis can be adopted using Eq. (7.1):

$$\sigma_m \approx 2\sigma_0 \left(\frac{c}{\rho}\right)^{1/2} \tag{7.1}$$

Where σ_0 the tensile stress, c is crack length, and ρ is the radius of the crack tip. The main conclusion of the analysis is that despite plastic deformation at the crack tip, failure is catastrophic [52].

Nanoindentation is another method for characterization of MEMS. One method used to calculate the fracture toughness of thin ceramic films is to measure the crack that initiates and grows upon loading (Fig. 7.9). By knowing the length of the crack, c, the peak load, P, the hardness, H, Young's modulus, E, and the type of indenter (represented by an empirical number α), the fracture toughness, K_{1C} can be calculated from Eq. (7.2) [53]:

$$K_{1C} \approx \alpha \left(\frac{E}{H}\right)^{1/2} \left(\frac{P}{c^{3/2}}\right) \tag{7.2}$$

Cracks are measured from the center of the indent to the end of the crack. An average can be obtained using all cracks initiating from the center of the indent. Cracking of topical brittle materials has been used to measure the creep properties of aluminum thin film sandwiched between the silicon nitride and silicon substrate. In this method, the topical Si_3N_4 layer is scratched, and the substrate is bent to cause

Fig. 7.9 Schematic of Ni-P cantilever beam and results of the test conducted [52]
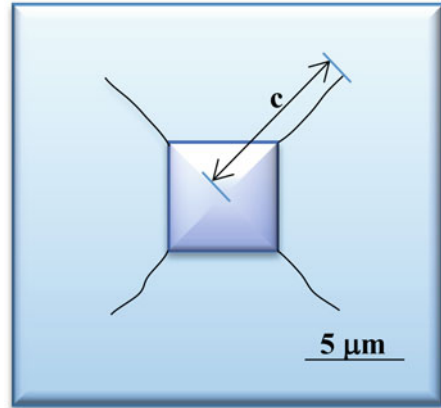
cracks grow. By measuring the crack growth rate, and the load applied in bending the viscosity of the aluminum underlayer is obtained [54].

On-the-Chip Tensile Testing: Efforts have been made to conduct tensile testing on the chips. This is achieved by building the actuator and sensor as well as the test specimen usually on one single chip, manufactured by semiconductor processing. A silicon-on-insulator (SOI) wafer composed of a 35-μm thick p-type device layer, a 2-μm thick buried oxide layer, and a 300-μm thick handle layer was used in one study to develop MEMS tension test device [55]. Deep reactive ion etching (DRIE) process using BOSCH recipe was carried out along with vapor HF acid etching to remove unwanted regions. These included boron oxide and buried silicon dioxide layers, respectively. Tensile test was conducted within scanning electron microscopy using capacitive comb drives to generate force. The force came from many sets of comb-structures electrostatically energized. Capacitive sensors were used for displacement and force measurements. The force is measured by multiplying the elongation of the specimen by the spring constant of the sample beam. Dimensioning of all suspension beams was performed with a scanning electron microscope, SEM. The resulting force and the resulting displacement could be measured with a resolution of 15 nN and 1 nm, respectively.

Nanowires of WC-Co were made using focused ion beam (FIB) microscopy from a bulk WC-Co specimen. The nanowires were then bonded onto the tensile test stage described above using W bonding inside the FIB microscope. Loading of the nanowires was conducted using the capacitive drives until the wires were fractured. Load and displacement was then carefully measured [55] (Fig. 7.9).

4 Modeling of MEMS

Modeling of MEMS devices has been carried out adopting approaches appropriate for their applications. In one such case [56], the nonlinear electromechanical

modeling of MEMS switches has been performed. The model takes into account force, damping effects, and position and calculates switching current. It also incorporates changes in the capacitance and voltage with time. The model predicts the switching time, current, velocity, and energy consumed in switching process. It was found out that taking into consideration the damping underneath MEMS bridge makes a large difference in energy and current predicted by this model vs those which ignore such damping.

Modeling software is available to simulate the function of MEMS both mechanically and electrically and characterize their behavior. One such model is COMSOL™ Multiphysics (COSMOL Inc., Burlington, MA). As an example, a silicon pressure sensor can be considered with a silicon membrane that has a cavity underneath isolating it electrically from the ground of the sensor. The pressure of the gas applied to the top pushes the membrane down, thus changing the capacitance between the membrane and the floor. The position of the membrane will be dictated by the pressure on top of it and can be determined from the value of the capacitance. To set up the model, a wizard function built into the simulation software lets the user select 3D space dimensioning. An electromechanics interface is then selected to combine solid mechanics and electrostatics with a moving mesh. This permits the modeling of the motion of the electrostatically actuated structures. There are various options to study, e.g., prestressed analysis, stationary, time dependent, and so on. The stationary study is used when variables do not change with time. It can compute static electric or magnetic fields and direct currents. In solid mechanics, it is used to calculate deformation stresses and strains. In fluid flow, it can determine steady flow and pressure fields. In chemical material transport, it is used to find the steady-state chemical compositions.

Parameters such as pressure, operating temperature, and die-bonding temperature (e.g., 20 kPa, 20 °C, and 70 °C, respectively) are imported from an external file along with geometry parameters. The next step would be to define boundaries (e.g., between the membrane, cavity underneath, and the silicon carbide die). An integrator is used to integrate the point of maximum deflection. Symmetry planes are then defined as *xz* and *yz*. There is a steel base to which the sensor is attached. That base will also be defined geometrically. It is the large disk base shown in Fig. 7.10. A cavity domain is then defined and its geometry specified. All components (ignoring the cavity) are defined to be linear elastic solids.

Physics aspects of the models are then specified. These include pressure felt by the membrane, voltage applied to the sensor, and the appropriate boundary conditions. In domains, linear elastic material is selected for all materials except for the vacuum cavity. In boundary selection, symmetry is selected. Symmetry planes are selected on the two vertical walls of the pie shown in Fig. 7.10. At this time, points are defined by selecting prescribed displacement. That prevents the motion of the bottom of the stainless steel base restricting deformation to that depth. In other words, prescribed displacement is constrained within the *z*-axis. Boundary load condition is defined next. This specifies the pressure that will be applied on the membrane shown by an arrow in Fig. 7.10. Load is defined as pressure and its value is set (e.g., at $P_0 = 20$ kPa). Mesh boundaries must be selected and the
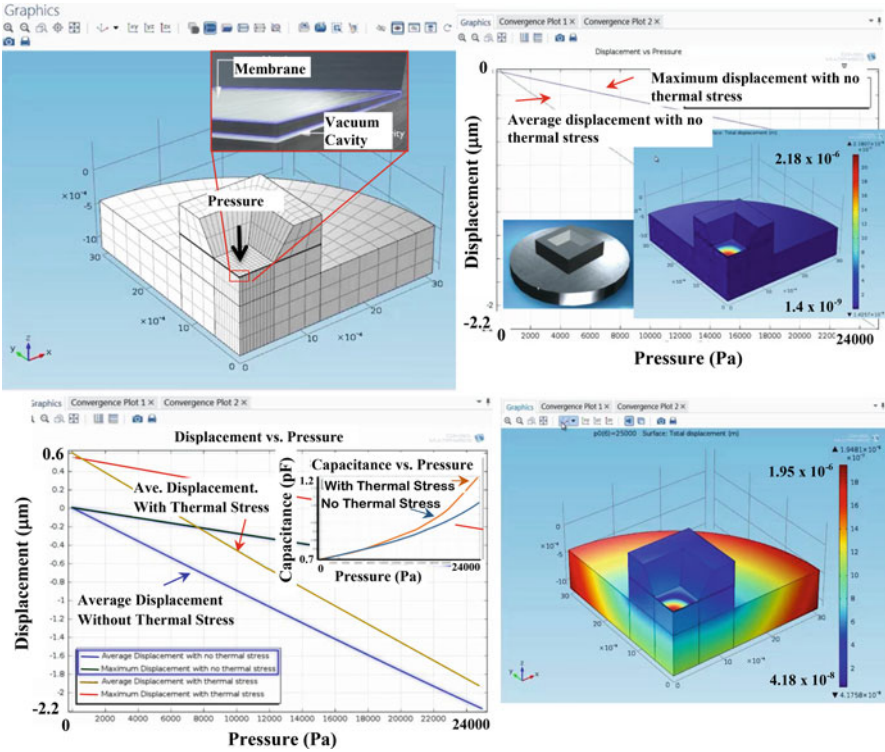
Fig. 7.10 Modeling of a silicon pressure sensor using COMSOL MultiPhysics (COSMOL Inc., Burlington, MA). *Top left*: Selection of the mesh with details of the center of the membrane, *Top right*: Full 3D schematic of the sensor, along with mapping of the displacement, *Bottom right*: Effect of thermal stresses on the variation of displacement at various locations; *Bottom left*: A plot showing the effect of thermal stresses on the variation of displacement and capacitance with pressure

boundaries next to the cavity. Electrostatic boundaries can be entered next. This tells the program about the amount of voltage applied to the underside of the membrane (e.g., 1 V). Following this, ground is selected in boundaries referring to the top of the silicon at the bottom of the cavity.

Materials selection will be performed after specifying boundaries. Silicon with values of 11.7 for relative permittivity, 170 GPa for elasticity modulus, 0.06 for Poisson's ratio, and 2330 kg/m^3 for density. For the cavity, a permittivity of 1 (e.g., vacuum) is entered. Lastly, for the base, Steel AISI 4340 is entered. Corresponding property values will be entered as 7850 kg/m^3 for density, 205 GPa for elasticity modulus, 0.28 for the Poisson's ratio, 1 for relative permittivity, 4.032×10^6 S/m for electrical conductivity, 12.3×10^{-6} for thermal expansion coefficient, and 44.5 W/m K for thermal conductivity. Creating the mesh will come next by selecting a mapped swept mesh with densest to be at the membrane. A size node is selected with a maximum element size of 50×10^{-5} m. The boundary

specified by the membrane is defined for this mesh. Mapped meshes are then selected for other parts. It is accomplished by the extension of the fine meshes of the central membrane to the perimeter of the disk. The final step is to sweep the mesh upward all the way to the top of the structure.

Once the geometry, properties, type of the materials, and the size and details of the mesh are configured, the simulation is performed. This involves studies on the application of a range of pressures and analyzing the response of the membrane. The study setting includes adding an auxiliary sweep of appropriate range of pressures with parameters that have a value list of range (0, 5000, 25,000). The computation function is then selected. This creates the map of displacements across the structure (shown as the insert in the pressure–displacement plot of Fig. 7.10). Plots of 1D and 2D can be created by specifying the average vertical displacement (e.g., in μm) and integration of the displacement (e.g., maximum displacement) shown in Fig. 7.10. With this model, the variation of capacitance vs. pressure can be obtained and plotted too. Interestingly, the effect of thermal expansion will also be the change in the capacitance due to thermal stresses. Of course, the coefficient of thermal expansion must also be entered into the model.

The results of the calculations are shown in terms of displacement vs pressure for both cases of presences and absence of thermal stresses. As seen from the plots (7.10), the presence of thermal stresses increase the displacement as well as the capacitance; however, the effect is more pronounced for capacitance at higher pressures. The effect of thermal stresses on the displacement are mapped in the bottom right. There are other simulation tools to characterize MEMS which also predict the behavior of the moving and stationary components and establish performance criteria for the actual device [57, 58].

Mathematical Modeling: Aside from software applications that allow modeling of MEMS components, mathematical modeling is another method adopted by many researchers. One such model is briefly described here [59]. It predicts the static output of the actuation of micromirrors. Consider a mirror plate driven by four Integrated Speed Control, ISC, actuators. The static response of the mirror is established based on the experimental height data obtained at different actuation voltages. The system is considered a 4-input, 4-output model. The actuation of each of the actuators would cause a height change to the corresponding side of the mirror. A 4×4 characteristic matrix relates the four actuation voltages as the input and the heights of the mirror as the output. Let character U denote the actuation voltage supplied to the device, the character d represent the resulting displacement, and subscripts specify each of the actuators and each side of the mirror plate. A voltage of 0–3 V is applied to the actual mirror, and the height change is measured by an autofocus microscope with a precision of 0.1 μm. Cubic functions are used to generate accurate fit for the height voltage curves.
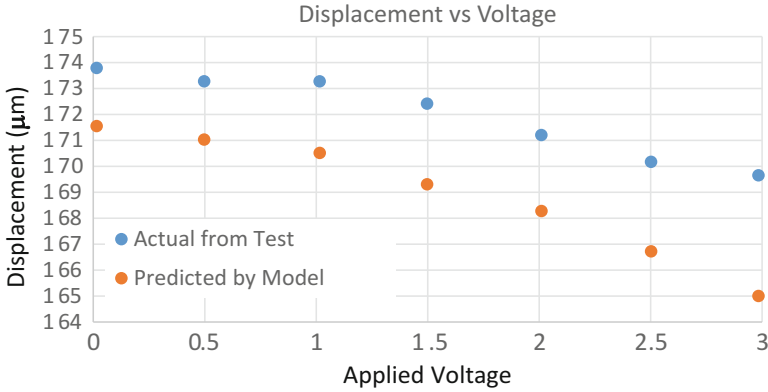
Fig. 7.11 Comparison between model predictions and actual test results [19]

$$\begin{pmatrix} f_{11} & f_{12} & f_{13} & f_{14} \\ f_{21} & f_{22} & f_{23} & f_{24} \\ f_{31} & f_{32} & f_{33} & f_{34} \\ f_{41} & f_{42} & f_{43} & f_{44} \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix}$$

Matrix elements are obtained from the coefficients of the four cubic functions established above. Linear superposition of the four sets of equations provide a height voltage matrix. Once the heights effected by applied voltages are calculated with the model, the tilt angles of the mirror plates are formulated as:

$$\theta_x = \arcsin \frac{|d_1 - d_3|}{W_x} \qquad \theta_y = \arcsin \frac{|d_2 - d_4|}{W_y}$$

Here, W_x and W_y denote the lengths of the sides of the mirror in x and y directions, and θ_x and θ_y represent the angles of the mirror plate in x and y axes. The accuracy of the model in predicting the height changes can be inferred from Fig. 7.11, where the predictions are plotted against measured data.

5   Future Trends

The market for MEMS devices is growing rapidly. As an example, the global MEMS market in the area of pressure sensors is going to surpass $4 billion by 2020 [60]. In 2015, the value of this market was about $2 billion. Most of the pressure sensor applications will be automotive, industrial, and healthcare use of which the most dominant is automotive. The total MEMS market was determined to have a value of $8 billion in 2005; however, this number increased to $14 billion in 2015, and furthermore up to $30 billion in 2020 [61]. Of this, about $3.2 billion was in sensors in 2015, and the rest was in inkjet print heads, catheters, and RF IC chips with embedded inductors. The growth of the MEMS market comes not

only from the increasing applications of current MEMS such as pressure sensors, triggers for airbags, and other automotive devices, RF MEMS, and miniaturized guidance systems, but also from new MEMS under development. These include disposable chips for performing assays on blood and tissue samples, integrated optical switching and processing chips, RF communication devices, and remote sensing products. Other areas suitable for new MEMS include structural health sensing, as well as detection, analysis, and mitigation of structural damage.

It is postulated that biosensing technologies will be as important in this half century as the microelectronics was to the past half century [62]. Expensive chemical and biological analyses will become inexpensively performed by on-site handheld biochip systems. These have better accuracy and results can be obtained in real time. The smaller size, reduced cost, shorter response time, smaller amount of fluid needed, reduced weight, and other advantages of biochips would radically transform the practice of clinical diagnostics. They make it more efficient and more affordable. It also makes sophisticated analytical techniques available in rural areas, especially developing countries that may lack conventional laboratories. The biochip market is expected to grow to nearly $9.6B in 2016 [62], which is a nearly threefold increase from 2010. It should be remembered that 40 % of the drugs being currently evaluated for clinical trials are transdermal or related to dermal systems. The value for the global transdermal patch market reached $2.8 billion in 2014 [63]. This signifies the importance of microneedles as potential careers for drug delivery of numerous macromolecular drugs for transdermal delivery. The growth in BioMEMS will include microneedles used for transdermal biosensing. The length scale is hundreds of microns and sense analytes such as glucose, biomarkers, and other substances of clinical interest. These microneedles will help develop biochips with end-effectors that interact with biological systems. They also minimize conventional injection-related infections, sample contamination, and analysis artifacts [62].

6 Summary and Concluding Remarks

A brief account of issues related to mechanical properties of MEMS was presented in this chapter. Microtesting techniques including microtensile and microfatigue testing along with the hardware were described with typical sample type, shape, and geometry, depicted with diagrams and images. Modeling MEMS using both computational and mathematical was illustrated by the aid of diagrams and plots. A short passage on future trends completed the chapter.

References

1. Pryputniewicz RJ (2012) Current trends and future directions in MEMS. Exp Mech 52(3): 289–303
2. Allameh SM, Lou J, Kavishe F, Buchheit T, Soboyejo WO (2004) An investigation of fatigue in LIGA Ni MEMS thin films. Mater Sci Eng A 371:256–266
3. Miller SL, Lavigne G, Rodgers MS, Sniegowski JJ, Waters JP, Mcwhorter PJ (1997) Routes to failure in rotating MEMS devices experiencing sliding friction. In: Proceedings of the SPIE—The International Society for Optical Engineering, vol 3224, pp 24–30
4. Allameh SM, Shrotriya P, Butterwick A, Brown S, Soboyejo WO (2003) Surface topography evolution and fatigue fracture in polysilicon MEMS structures. J Microelectromech Syst 12:313–324. doi:10.1109/JMEMS.2003.809957
5. Muhlstein CL et al (2001) High-cycle fatigue and durability of polycrystalline silicon thin films in ambient air. Sens Actuators A Phys 94(3):177–188
6. Brown SB, Van Arsdell W, Muhlstein CL (1997) Materials reliability in MEMS devices. Presented in proceedings of international solid state sensors and actuators conference (Transducers'97), IEEE, New York, vol 1, pp 591–593
7. Brown SB, Jansen E (1996) Reliability and long term stability of MEMS. Advanced applications of lasers in materials processing/broadband optical networks/smart pixels/optical MEMs and their applications. In: IEEE/LEOS 1996 summer topical meetings, Keystone, CO, 5–9 Aug 1996, pp 9–10. doi:10.1109/LEOSST.1996.540771
8. Yu HH, Suo Z (2000) Stress-dependent surface reactions and implications for a stress measurement technique. J Appl Phys 87:1211–1218
9. Liang J, Suo Z (2001) Stress-assisted reaction at a solid-fluid interface. Interface Sci 9:93–104
10. Yang WH, Srolovitz DJ (1994) Surface morphology evolution in stressed solids: surface diffusion controlled crack initiation. J Mech Phys Solids 42(10):1551–1574
11. Tanner DM (2000) Reliability of surface micromachined microelectromechanical actuators. In: Proceedings of 22nd international conference on microelectronics. Proceedings, 2000, IEEE, Piscataway, NJ, pp 97–104. doi:10.1109/ICMEL.2000.840535
12. Smith NF, Eaton WP, Tanner DM, Allen JJ (1999) Development of characterization tools for reliability testing of microelectromechanical system actuators. In: MEMS reliability for critical and space applications. Proceedings of SPIE—The International Society for Optical Engineering, vol 3880, pp 156–164
13. Tanner D, Miller W, Peterson K, Dugger M, Eaton W, Irwin L, Senft D, Smith N, Tangyunyong P, Miller S (1999) Frequency dependence of the lifetime of a surface micromachined microengine driving a load. Microelectron Reliab 39:401–414
14. Tanner DM, Walraven JA, Helgesen K, Irwin LW, Brown F, Smith NF, Masters N (2000) MEMS reliability in shock environments. Presented in annual proceedings—reliability physics (symposium), IEEE, Piscataway, NJ, pp 129–138
15. Wagner U, Franz J, Schweiker M, Bernhard W, Muller-Fiedler R, Michel B, Paul O (2001) Mechanical reliability of MEMS-structures under shock load. Microelectron Reliab 41(9/10):1657–1662
16. Miller SL, Rodgers MS, LaVigne G, Sniegowski JJ, Clews P, Tanner DM, Peterson KA (1998) Failure modes in surface micromachined microelectromechanical actuators. In: 1998 IEEE international reliability physics symposium proceedings, IRPS 1998, 31 Mar–2 Apr 1998, pp 17–25
17. Tanner DM, Walraven JA, Irwin LW, Dugger MT, Smith NF, Eaton WP, Miller WM, Miller SL (1999) Effect of humidity on the reliability of a surface micromachined microengine. Presented in annual proceedings—reliability physics (symposium), IEEE, Piscataway, NJ, pp 189–197. doi:10.1109/RELPHY.1999.761611
18. Takashima K, Higo Y, Sugiura S, Shimojo M (2001) Fatigue crack growth behavior of micro-sized specimens prepared from an electroless plated Ni-P amorphous alloy thin film. Mater Trans 42(1):68–73. doi:10.2320/matertrans.42.68

19. Zhang GP, Takashima K, Shimojo M, Iligo Y (2000) Fatigue behavior of microsized austenitic stainless steel specimens. Mater Lett 57:1555–1560. doi:10.1016/S0167-577X(02)01023-6

20. Tsuchiya T, Sakata J, Taga Y (1998) Tensile strength and fracture toughness of surface micromachined polycrystalline silicon thin films prepared under various conditions. In: Cammarata RC, Nastasi M, Busso EP, Oliver WC (eds) Presented in thin-films-stresses and mechanical properties VII. Symposium, 1998, Mater. Res. Soc., Warrendale, PA, pp 285–290. doi:10.1557/PROC-505-285

21. Tanner DM, Walraven JA, Helgesen KS, Irwin LW, Gregory DL, Stake JR, Smith NF (2000) MEMS reliability in a vibration environment. Presented in annual proceedings—reliability physics (symposium), IEEE, Piscataway, NJ, pp 139–145. doi:10.1109/RELPHY.2000.843904

22. White CD, Shea HR, Cameron KK, Pardo F, Bolle CA, Aksyuk VA, Arney S (2000) Electrical and environmental reliability characterization of surface-micromachined MEMS polysilicon test structures. In: Proceedings of the SPIE—The International Society for Optical Engineering, vol 4180, pp 91–95. doi:10.1117/12.395697

23. Marxer C, Gretillat MA, De Rooij NF, Batting R, Anthametten O, Valk B, Vogel P (1997) Reliability considerations for electrostatic polysilicon actuators using as an example the REMO component. Sens Actuators 61(1–3):449–454. doi:10.1016/S0924-4247(97)80304-4

24. Patton ST, Cowan WD, Zabinski JS (1999). Performance and reliability of a new MEMS electrostatic lateral output motor. Presented in annual proceedings-reliability physics (symposium), IEEE, Piscataway, NJ, pp 179–188. doi:10.1109/RELPHY.1999.761610

25. Lafontan X, Pressecq F, Perez G, Dufaza C, Karam JM (2001) Physical and reliability issues in MEMS microrelays with gold contacts. In: Proceedings of the SPIE—The International Society for Optical Engineering, vol 4558, pp 11–21

26. Smith BK, Brown CD, Lavigne G, Sniegowski JJ (1998) Thin Teflon-like films for MEMS: film properties and reliability studies. In: Proceedings of the SPIE—The International Society for Optical Engineering, vol 3511, pp 114–125. doi:10.1117/12.324289

27. Bahr DF, Merlino JC, Banerjee P, Yip CM, Bandyopadhyay A (1999) Reliability and properties of PZT thin films for MEMS applications. In: Heuer AH, Jacobs SJ (eds) Presented in materials science of microelectromechanical systems (MEMS) devices, Mater. Res. Soc., Warrendale, PA, vol 546, pp 153–158. doi:10.1557/PROC-546-153

28. Renaud M, Fujita T, Goedbloed M, de Nooijer C, van Schaijk R (2014) Improved mechanical reliability of MEMS piezoelectric vibration energy harvesters for automotive applications. J Micromech Microeng 25(10):568–571. doi:10.1109/MEMSYS.2014.6765704

29. Soboyejo WO, Allameh S, Gally B, Brown S, Freeman D, Evans AG (2000) Reliability of mechatronic silicon MEMS structures. In: Ume C (ed) Proceedings of the 7th mechatronics forum international conference, Elsevier, Atlanta, GA. Paper no. 41.03

30. Bhushan B (2001) Modern tribology handbook, vol 1. Principles of tribology; vol 2. Materials, coatings, and industrial applications. CRC, Boca Raton, FL

31. Dhennin J, Lellouchi D, Pressecq F (2015) How to evaluate the reliability of MEMS devices without standards. In: Proceedings of 2015 symposium on design, test, integration and packaging of MEMS/MOEMS, Montpellier, France, 27–30 Apr 2015, pp 254–257. doi:10.1109/DTIP.2015.7161015

32. Pathak R, Joshi S (2009) Reliability modeling and optimization of MEMS elements in various devices using multi-scale concepts. In: Innovative technologies in intelligent systems and industrial applications, CITISIA, pp 332–337. doi:10.1109/CITISIA.2009.5224187

33. Fang X-W, Huang Q-A, Tang J-Y (2004) Modeling of MEMS reliability in shock environments. In: Proceedings of the solid-state and integrated circuits technology, vol 2, pp 860–863. doi:10.1109/ICSICT.2004.1436643

34. Maier-Schneider D, Maibach J, Obermeier E (1995) A new analytical solution for the load-deflection of square membranes. J Microelectromech Syst 4(4):238–241. doi:10.1109/84.475551

35. Tanner DM, Smith NF, Irwin LW, Eaton WP, Helgesen K, Clement JJ, Miller WM, Walraven JA, Peterson KA, Tangyunyong P, Dugger MT, Miller SL (2000) MEMS reliability: infrastructure, test structures, experiments, and failure modes. In: Sandia National Laboratories,

report, SAND2000-0091, unlimited release, pp 1–171. http://www.sandia.gov/mems/_assets/documents/bibliography/3_13MEMS.pdf

36. Que L, Park J, Gianchandani Y (1999) Bent-beam electro-thermal actuators for high force applications. In: IEEE international conference on micro electro mechanical systems, Orlando, FL, Jan 1999, pp 552–557
37. Lee T, Seo Y, Whang K, Choi D (2006) Study on the lateral actuator with actuation range amplifying structure. Key Eng Mater 326–328:289–292
38. Trolier-McKinstry S, Muralt P (2004) Thin film piezoelectrics for MEMS. J Electroceram 12(1):7–17
39. Gee D, Currano L (2007) Fabrication and testing of a novel MEMS rotational thermal actuator. US Army Research Lab., Adelphi, MD report no. ARL-TR-4315, Nov 2007, pp 1–22
40. Sharpe WN Jr, McAleavey A (1998) Tensile properties of LIGA nickel. In: Proceedings of SPIE—The International Society for Optical Engineering, vol 3512, pp 130–137
41. Yang Y, Allameh SM, Lou J, Imasogie B, Boyce BL, Soboyejo WO (2007) Fatigue of LIGA Ni micro-electro-mechanical systems thin films. Metall Mater Trans A 38:2340–2348
42. Wardlow J, Allameh SM (2015) On the micromechanical characterization of metallic MEMS by a hybrid microtester. In: Proceedings of 2015 ASME international mechanical engineering congress and exposition, Houston, TX, 13–19 Nov 2015
43. Zhou J, Allameh SM, Soboyejo WO (2005) Microscale testing of the strut in open cell aluminum foams. J Mater Sci 40:429–439
44. Arsène MA, Savastano H Jr, Allameh SM, Ghavami K, Soboyejo WO (2003) Cementitious composites reinforced with vegetable fibers. In: Proceedings of IAC NOCMAT 2003, first inter American conference on non-conventional materials and technologies in the eco-construction and infrastructure-IAC NOCMAT 2003, João Pessoa, Brazil
45. Allameh SM, Sadat Hossieny M, Rajai M (2005) Development of microtesting systems: I. Tensile testing of metallic microsamples. In: Proceedings of 2005 ASEE annual meeting and conference, Portland, OR, No. 2005-1010
46. Allameh SM, Gally B, Brown S, Soboyejo WO (2001) Surface topology and fatigue in Si MEMS structures. In: Muhlstein C, Brown S (eds) Mechanical properties of structural films, STP 1413. American Society for Testing and Materials, West Conshohocken, PA, pp 3–16
47. Freeman DM, Aranyosi AJ, Gordon MJ, Hong SS (1998) Multidimensional motion analysis of MEMS using computer microvision. In: Proceedings of the solid-state sensor and actuator workshop, Hilton Head Island, SC, June 1998, pp 150–155
48. Quentin Davis C, Freeman DM (1998) Using a light microscope to measure motions with nanometer accuracy. Opt Eng 37:1299–1304
49. Ritchie RO (1988) Mechanisms of fatigue crack propagation in metals, ceramics and composites: role of crack tip shielding. Mater Sci Eng A A103:15–28
50. Shrotriya P, Allameh S, Butterwick A, Brown S, Soboyejo WO (2002) Mechanisms of fatigue in polysilicon MEMS structures. Mater Res Soc Symp Proc 687(B2.3):29–34
51. Allameh SM, Gally B, Brown S, Soboyejo WO (2001) On the evolution of surface morphology of polysilicon MEMS structures during fatigue. In: Kahn H et al (eds) MRS fall meeting, symposium EE: materials science of microelectromechanical system (MEMS) devices III, Boston, MA, 2000, MRS proceeding, paper, vol 657, pp EE2.3.1–EE2.3.6
52. Li X, Bhushan B, Takashima K, Baek CW, Kim YK (1997) Mechanical characterization of micro/nanoscale structures for MEMS/NEMS applications using nanoindentation techniques. Ultramicroscopy 97(1–4):481–494
53. Lawn BR, Evans AG, Marshall DB (1980) Elastic/plastic indentation damage in ceramics: the median/radial crack system. J Am Ceram Soc 63:574–581
54. Allameh SM, Suo Z, Soboyejo W (2007) Creep of Al underlayer determined by channel cracking of topical Si_3N_4 film. J Mater Manuf Process 22:170–174
55. Namazu T, Morikaku T, Akamine H, Fujii T, Kuroda K, Takami Y (2015) Mechanical reliability of FIB-fabricated WC–Co cemented carbide nanowires evaluated by MEMS tensile testing. Eng Fract Mech 150:126–134. doi:10.1016/j.engfracmech.2015.07.007

56. Muldavin JB, Rebeiz GM (2001) Nonlinear electro-mechanical modeling of MEMS switches. In: Proceedings of conference on microwave symposium digest, 2001 IEEE MTT-S international, 20–24 May 2001, Phoenix, AZ, vol 3, pp 2119–2122. doi:10.1109/MWSYM.2001.967332

57. Shen MC, Nichols J, Garson C, Mills I, Matar M, Fewell J, Pant K, Prabhakarpandian B (2015) Synthetic tumor networks for screening drug delivery systems. J Control Release 201:49–55. doi:10.1016/j.jconrel.2015.01.018. CFDRC (2015) CFD-ACE + Multiphysics software, http://www.cfdrc.com

58. Song H, Wang Y, Garson C, Pant K (2015) Concurrent DNA preconcentration and separation in bipolar electrode-based microfluidic device. Anal Methods 7:1273–1279. doi:10.1039/C4AY01858C

59. Zhang H, Xu D, Zhang X, Chen Q, Xie H, Li S (2015) Model-based angular scan error correction of an electrothermally-actuated MEMS mirror. Sensors 15:30991–31004. doi:10.3390/s151229840

60. Technavio Research (2016) Global MEMS pressure sensor market to exceed USD 4 billion by 2020. Business Wire, London, Accession Number: bizwire.c66356480. 19 Jan 2016

61. Mounier E, Troadec C, Girardin G, de Charentenay Y (2015) Status of MEMS industry. Yole Development Report (May 2015), pp 40–61. Cited in Market & Technology on i-Micronews.http://www.i-micronews.com/component/hikashop/product/status-of-the-mems-industry-2015.html

62. Ventrelli L, Marsilio Strambini L, Barillaro G (2015) Microneedles for transdermal biosensing: current picture and future direction. Adv Healthc Mater 4(17):2606–2640. doi:10.1002/adhm.201500450

63. Akhtar N (2014) Microneedles: an innovative approach to transdermal delivery—a review. Int J Pharm Pharm Sci 6(4):18–25

Chapter 8
Basic Theory and Modelling of Marmot-Like Robot for Mine Safety Detection and Rescuing

Dan Zhang and Bin Wei

Abstract Mine rescue robots are a kind of intelligent robot which can work in a complex and dangerous underground environment and be used for detecting human beings. Mine rescue robot to be developed in this chapter includes a system with certain function of drilling and excavation, a sensing system, and an external support system of wireless communication. The kinematic modelling of the head section of a type of rescue robot is presented, and the maximum stiffness, minimum stiffness, and global stiffness of the head section of a type of rescue robot is analyzed.

1 Introduction

Recently, biologically inspired manipulators have attracted many researchers' interests. Many universities in the United States have manufactured prototypes of the creeper models and investigated other biologically inspired manipulators. Furthermore, stiffness is a one of the vital elements to assess robot performance. Stiffness analysis on robotic manipulators has been a popular topic among numerous researchers over the past decade. For examples, Tsai and Joshi [1, 2] investigated the stiffness properties of 3-DOF position mechanisms for use in the hybrid kinematic machines. Zhang and Gosselin [3] have studied the stiffness and compliance of N-DOF parallel manipulators while considering the lumped kinetostatic model accounting for joint and link compliance. Other scholars [4] have researched the stiffness of 3-PUU parallel robot, etc.

D. Zhang
Department of Mechanical Engineering, Lassonde School of Engineering,
York University, 4700 Keele Street, Toronto, ON, Canada M3J 1P3
e-mail: dzhang99@yorku.ca

B. Wei (✉)
Faculty of Engineering and Applied Science, University of Ontario Institute of Technology,
2000 Simcoe Street North, Oshawa, ON, Canada L1H 7K4
e-mail: Bin.Wei@uoit.ca

When there is an accident at a coal mine, the rescue robot can obtain environmental detecting parameters and hazard information, and discover these survivors information and transfer it to the ground, so as to minimize the loss of lives and property. A groundhog-like detecting and rescuing robot system has been developed [5] aiming at detecting the location of unreachable accident due to a tunnel being blocked in a coal mine tragedy. This system is capable of crawling or drilling through collapsed objects to enter the accident location, to detect flammable and explosive gas density, and to send the measured signals back to the ground via a wireless network. To the best of the authors' knowledge, this kind of robot, driven by both legs and wheels, has never been reported. This groundhog-like detecting and rescuing robot system, including a parallel-robot-based drill, a robot body and its control unit, can be used to detect gas density of an accident environment to alert further dangers, as well as to communicate with the rescue team via a wireless network. Based on the bionics principle of groundhog and parallel robot mechanisms, a robot drilling head which can rotate 360° with solid stiffness was successfully developed [6]. The multi-joint body can wriggle forward in narrow space, and digging legs can push the coal backwards. The kinematic modelling of the head section of a type of rescue robot is presented, and the maximum stiffness, minimum stiffness, and global stiffness of the head section of a type of rescue robot is analyzed. For the gas detection and alarm system, an early warning esplosion model was developed based on flammable gas density detection by means of introducing cold plasma for the catalysis element with an enlarged surface area.

2 System Design and Analysis

2.1 Wheeled and Legged Robot

The robot proposed here consists of a rigid body and six semicircular elastic legs, which not only can achieve forward, backward, and turning movements by alternating tripod gait, but also can achieve more flexible and complex forms of exercise such as jumping and climbing stairs. Each leg is connected to the body via a revolute joint. By symmetrical design of the structure, the robot can move without considering the movement directions, and the motion adaptability in the complex environment is greatly improved. Performance parameters and the prototype are illustrated below (Fig. 8.1).

Dimensions: 48 cm × 25 cm × 12 cm
Power: ≤ 72 W, Weight: 10.5 kg
Leg length: 29 cm, velocity: ≤ 0.45 cm/s

2.2 Crawling Robot

A crawling robot body consists of the following five sub-modules: a head and neck module used to dig the earth, a front support module, an intermediate creep-feed

Fig. 8.1 Prototype of the robot
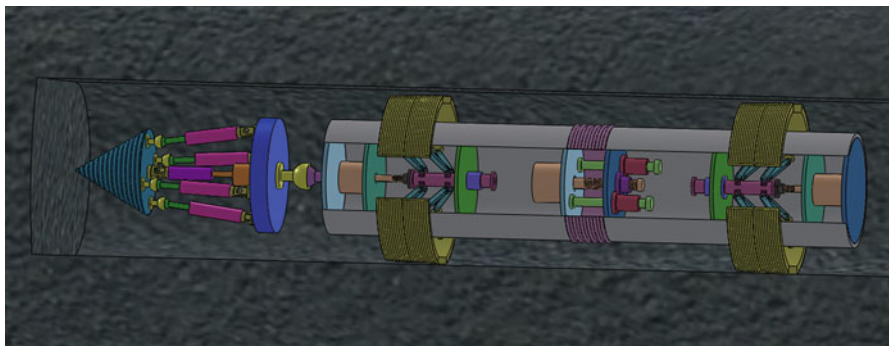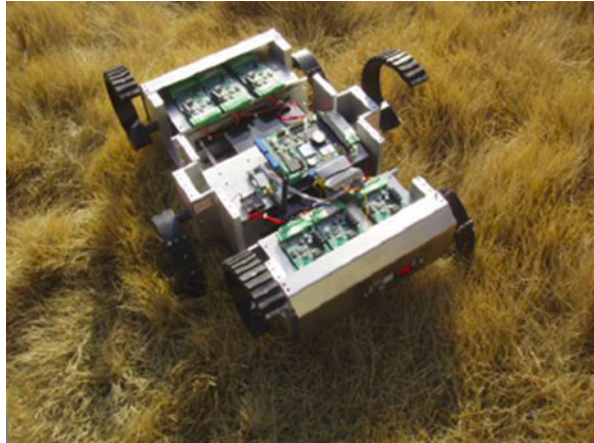
Fig. 8.2 Schematic of crawling robot

module, a back support module, and a commutation and maintenance module. The
robot can achieve squirm movement by coordination of radial swelling and axial
elongation contraction, as well as realizing turning function within the roadway
by resorting to a commutation module. The performance parameters are listed as
follows and the prototype is shown in Fig. 8.2.

Radial dimension: $130 \sim 160$ mm, total length: $0.9 \sim 1.1$ m, squirm feed rate:
10 mm/s, total weight: ≤ 10 kg.

2.3 Marmot-Like Robot

Based on the above modules, and aimed at detecting the scenario of unreachable
accident location due to tunnel block in a coalmine tragedy, and based on the bionics
principle of groundhog and parallel robot mechanism, three types of marmot-like
robot system are developed as shown in Fig. 8.3. The robot consists of legs, a
waist, and a head. The control system, which consists of a PC monitoring system,
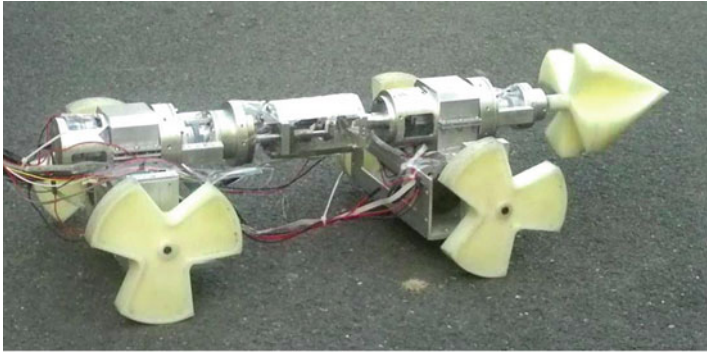
Type I
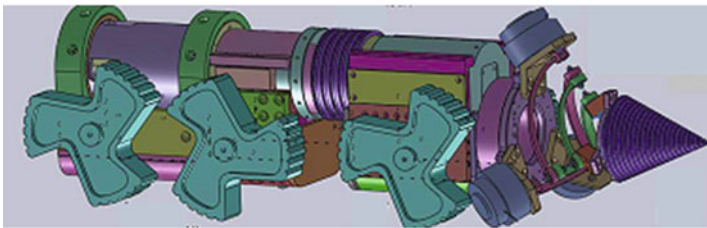
Type II
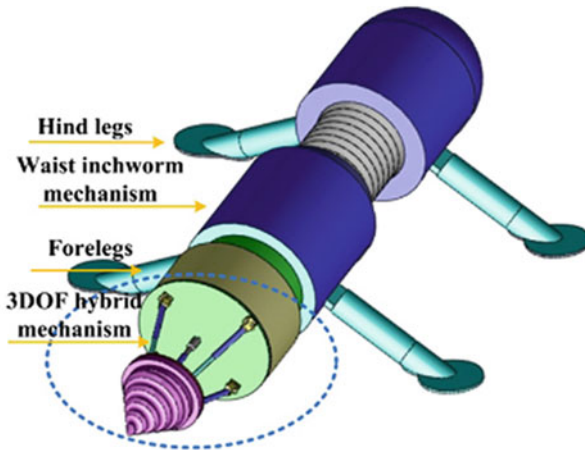
Type III

Fig. 8.3 Prototype of Marmot-like robot

a lower computer control system, a motor control system, and various types of sensors and communication interfaces, is based on the hierarchical structure. The spring-mass inverted pendulum (SLIP) model can accurately describe the dynamics of walking animals, which can be used for a similar structure robot dynamic control. The system alternates between a support phase and an aerial phase while SLIP

Fig. 8.4 BSLIP model
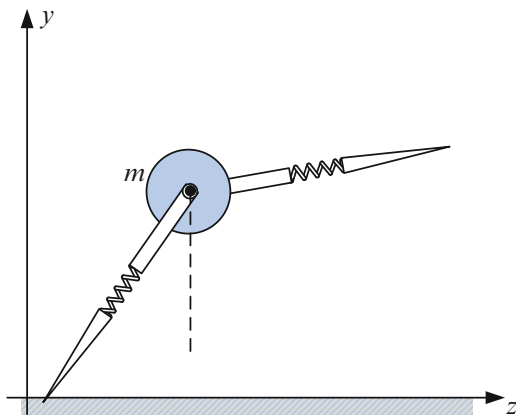
is in motion. The support phase can be divided into a compression phase and a decompression phase, and the aerial phase is divided into rising phase and decline phase. The critical conditions for conversion of the four subphases are as follows:

Touchdown transient: instant status of the legs from swinging in the air to touching the ground.
Bottom transient: instant status of the maximum amount of leg compression in the support phase.
Lifting transient: instant status of the feet off the ground.
Vertex transient: instant status of the body reaching the maximum height in the aerial phase.

By adding another leg in the SLIP, a two-legged spring-mass pendulum model (BSLIP) is derived. A BSLIP model is similar to the two-legged hexapod robot, so its control is similar to the hexapod robot (Fig. 8.4).

The robot dynamic equation can be derived based on the BSLIP model. The motion model and controller are developed for the BSLIP model based the hexapod robot. Their main functions include the following:

1. Applying a rotation tripod gait signal to identify BSLIP system triangle
2. Selecting an appropriate virtual foot to obtain the amount of control for the BSLIP system when the legs touch the ground.
3. Calculating appropriate driving torque to drive the center of mass of the robot system
4. Pendulum dynamic control.

Tests show that the robot can achieve line and turning motion on the ground, and the maximum speed can be up to 0.45 m/s. However, there will be gait imbalances when the robot conducts high speed walking for a long period of time, and the only way to achieve balance is to reset the robot. The possible reason is that lack of precision in machining the parts, such as the inconsistency of center distance of gears, and gear meshing is not ideal.

# 3 Stiffness Analysis for the Type III Marmot-Like Rescue Robot

Stiffness is a measurement of the ability of a body to withstand the action of deformation because of action of external forces. The stiffness of a mechanism at a given point within the workspace can be evaluated by the stiffness matrix. The stiffness matrix relates the forces/torques that are applied at the end-effector in Cartesian space to its Cartesian displacements. Stiffness can be a critical element in many scenarios due to the fact that it relates to accurate positioning and other dynamic performances. Here, we first investigated the maximum and minimum stiffness of the head section of the type III Marmot-like rescue robot, which has 3 degrees of freedom, i.e., translation along z axis and rotations about x axis and y axis; secondly, we investigated the global stiffness and its optimization in order to obtain the optimal stiffness in each direction.

3.1 Geometric Modeling

When the accident occurred at the coal mine, the rescue robot can obtain environmental detecting parameters and hazard information, and discover these survivors information and transferred it to the ground, and to help emergency rescue teams to have their searching and rescue planned, so as to minimize the loss of lives and property. The type III rescue robot includes a robot drilling head which can rotate with solid stiffness, a multi-joint body which can wriggle forward in narrow space, and digging legs which can push the coal backwards. This system is capable of crawling or drilling through the collapsed objects to enter the accident spot, to detect the flammable gas density, and to send the measured signals back to the ground via wireless network. This groundhog-like detecting and rescuing robot system can be used to detect gas density of accident environment and alert further dangers, as well as to communicate with the rescuing team via wireless network.

As shown in Fig. 8.5, the head section mechanism has four identical variable-length legs and one passive constraining leg that connect the fixed base to a moving platform. In this manipulator, each of the four identical legs consists of a universal joint, a moving link, an actuated prismatic joint, another moving link, and a spherical joint attached to the platform. The middle chain that connects the base center to the platform center is a passive constraining leg, and it consists of a prismatic joint attached to the base, a moving link, and a universal joint attached to the platform. The last leg is used to force the motion of the platform to three degrees of freedom. For the purpose of analysis, a fixed reference frame (XYZ) is attached to the base at point O, located at the center of the circumferential circle of the base. For each leg, another coordinate system (uvw)$_i$ is attached to the fixed base at B_i, such that u_i is in direction of b_i and at angle φ_i from X-axis as shown in Fig. 8.1. A moving frame (xyz) is attached to the moving platform at point O_e, located at the center of the circumferential circle of the moving platform. The angle between a_i and x-axis is defined φ_i'. Here, we assume $\varphi_i' = \varphi_i$.

Fig. 8.5 Schematic representation of the head section mechanism
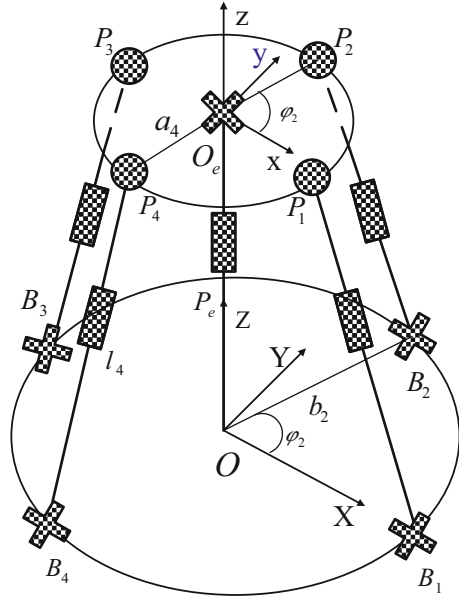
3.2 Inverse Kinematics

As shown in Fig. 8.5, all vectors are expressed in fixed reference frame as follows:

$$b_i = \begin{bmatrix} r \cos \varphi_i & r \sin \varphi_i & 0 \end{bmatrix}^T \tag{8.1}$$

$$a_i = Ra'_i \tag{8.2}$$

Here,

$$a'_i = \begin{bmatrix} c \cos \varphi'_i & c \sin \varphi'_i & 0 \end{bmatrix}^T \tag{8.3}$$

which r and c are the radii of the base and moving platform. Moreover, R is the rotation matrix of the moving platform with respect to the base and can be written as,

$$R = \begin{bmatrix} \cos \psi & 0 & \sin \psi \\ \sin \varphi \sin \psi & \cos \varphi & -\sin \varphi \cos \psi \\ -\cos \varphi \sin \psi & \sin \varphi & \cos \varphi \cos \psi \end{bmatrix} \tag{8.4}$$

The position vector of point O_e with respect to origin of the fixed coordinate frame can be written as $\begin{bmatrix} 0 & 0 & h \end{bmatrix}^T$.

The vector l_i can be written as,

$$l_i = p_e + a_i - b_i \tag{8.5}$$

The length of the ith leg is written as,

$$q_i^2 = l_i^T l_i \tag{8.6}$$

The Jacobian matrix can be determined by time-differentiating equation (8.6) as follows:

$$B\dot{q} = At \tag{8.7}$$

$$J = B^{-1}A \tag{8.8}$$

where A and B are defined for the manipulator as [7]

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}, \quad B = \begin{bmatrix} q_1 & 0 & 0 & 0 \\ 0 & q_2 & 0 & 0 \\ 0 & 0 & q_3 & 0 \\ 0 & 0 & 0 & q_4 \end{bmatrix}, \quad t = \begin{bmatrix} \dot{\varphi} \\ \dot{\psi} \\ \dot{h} \end{bmatrix}, \cdots \dot{q} = \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \\ \dot{q}_4 \end{bmatrix}$$

where the elements of A can be written as

$$a_{i1} = -rc \, \sin \, \varphi_i \cos \, \varphi_i \cos \, \varphi \, \sin \, \psi + rc \, \sin \, \varphi \, \sin^2 \varphi_i + hc \, \cos \, \varphi_i \sin \, \varphi \, \sin \, \psi$$
$$+ hc \, \sin \, \varphi_i \cos \, \varphi$$
$$a_{i2} = rc \, \cos^2 \varphi_i \sin \, \psi - rc \, \sin \, \varphi_i \cos \, \varphi_i \sin \, \varphi \, \cos \, \psi - hc \, \cos \, \varphi_i \cos \, \varphi \, \cos \, \psi$$
$$a_{i3} = h - c \, \cos \, \varphi_i \sin \, \psi \, \cos \, \varphi + c \, \sin \, \varphi_i \sin \, \varphi \quad i = 1, 2, 3, 4$$

3.3 Maximum and Minimum Stiffness

When a mechanism conducts a mission, the end-effector applies force to the environment. The reaction force will make the end-effector to be deflected away from its desired location. The deflection amount can be written as a function of applied force and the manipulator stiffness. So the stiffness of a manipulator has an impact on its accuracy of positioning. The stiffness of a mechanism is depending on several factors, includes the size and material of the links, mechanical transmission mechanisms, actuators, and control system, etc. In this chapter, we assume the main source of the compliance come from the compliance of the actuators.

Let F denote end-effector output force vector and Δx represent the end-effector displacement vector. It can be shown that:

$$F = K\Delta x \qquad (8.9)$$

where

$$K = J^T K_J J \qquad (8.10)$$

is known as the stiffness matrix, and K_J is a 4×4 diagonal matrix in which each non-zero diagonal element, k_i, represents the stiffness constant of the ith joint actuator. Furthermore, if $k_1 = k_2 = k_3 = k_4 = k$, the above equation reduces to

$$K = kJ^T J \qquad (8.11)$$

We observe from Eq. (8.10) that K is a 3×3 symmetric and position-dependent matrix. In addition, the deflection is relying on output force direction, in other words, the stiffness of a manipulator is direction dependent. Let λ_i be an eigenvalue and f_i be the corresponding eigenvector of the stiffness matrix K at a designated platform location. Then, λ_i is the stiffness in the corresponding eigenvector direction. It follows that if $\lambda_{\min}$ is the minimum eigenvalue and $\lambda_{\max}$ is the maximum eigenvalue, then the minimum stiffness takes place in the $f_{\min}$ direction and maximum stiffness takes place in the $f_{\max}$ direction.

The stiffness constant is taken to be 1000 N/m for a liner actuator. The maximum and minimum stiffness values at each point within the workspace of the head section of the type III rescue robot have been computed. The maximum stiffness mapping at the elevation of $z = 0.17$ m as a case for the manipulator is shown in Fig. 8.6. The minimum stiffness mapping at the elevation of $z = 0.17$ as a case for the manipulator is shown in Fig. 8.7. Figure 8.6 shows that the highest stiffness value occurs when ψ and φ are equal to zero. Figure 8.7 shows that the lowest stiffness value occurs near the manipulator workspace boundary; this is because a manipulator becomes singular when it approaches the workspace boundary.

3.4   Stiffness Optimization of 4UPS-PU Manipulator

3.4.1   Global Stiffness of the Manipulator

As stated in Eq. (8.11), the stiffness of the manipulator is expressed by a 3×3 matrix. The diagonal elements of the matrix express the pure stiffness of the mechanism in each direction. In order to obtain the optimal stiffness in each direction, one can write an objective function (8.12), with stiffness element to maximize. The global stiffness K_{global} is defined as follows:

$$K_{\text{global}} = \eta_1 k_{11} + \eta_2 k_{22} + \eta_3 k_{33} \qquad (8.12)$$
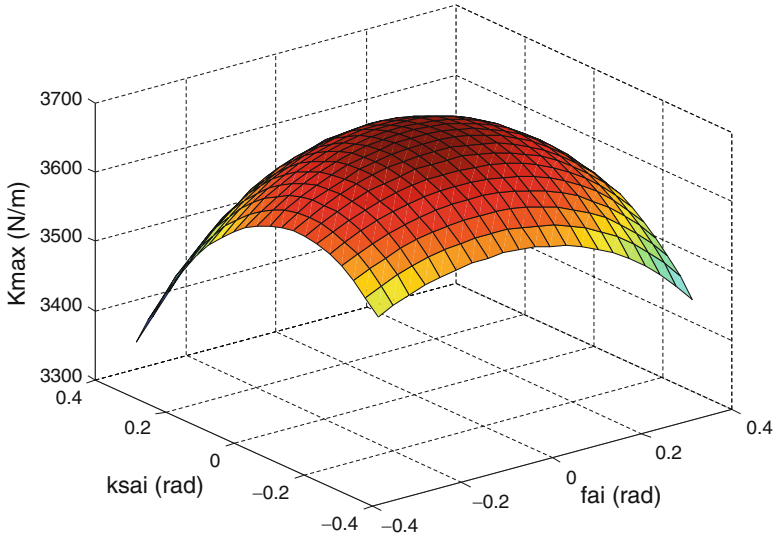
Fig. 8.6 Maximum stiffness of the head section of the type III rescue robot
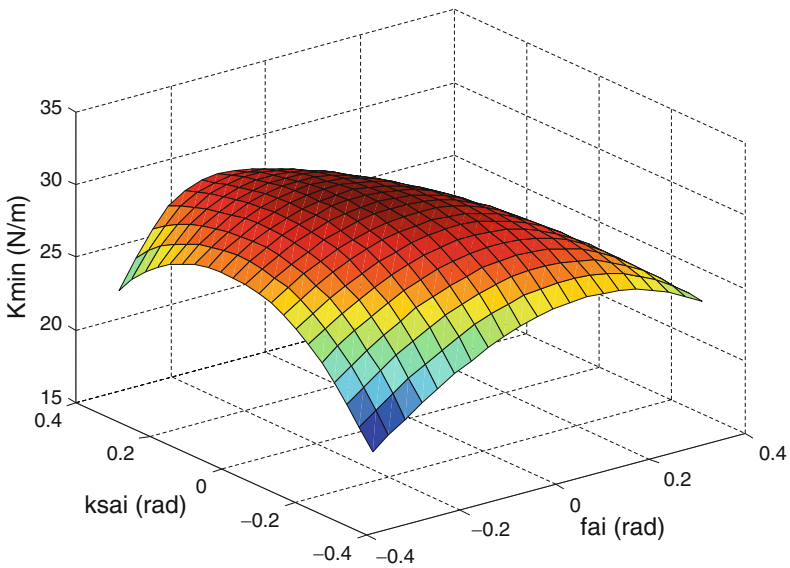
Fig. 8.7 Minimum stiffness of the head section of the type III rescue robot

where k_{ii} ($i = 1, 2, 3$) is the diagonal element of the stiffness matrix of the mechanism. η_i is the weight factor for each directional stiffness, which evaluates the priority of the stiffness in this direction. This would maximize the sum of the diagonal element which is global stiffness. Even though we couldn't get each diagonal element individually maximized, each stiffness can be optimized by distributing the weighting factors.
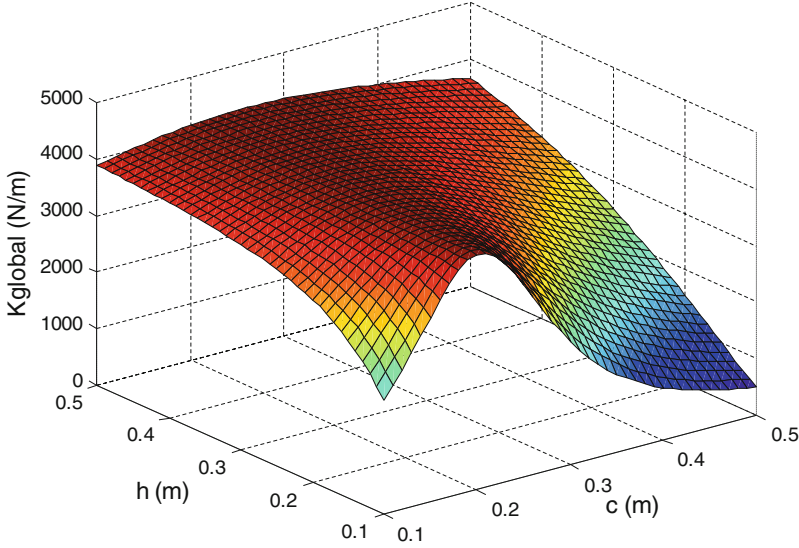
Fig. 8.8 K_{global} distribution when $r = 0.2$ m, $\varphi = 0°$, $\psi = 0°$

For the purpose of numerical analysis first, we assume $r = 0.2$ m, $\varphi = 0$, $\psi = 0$, $\eta_i = 1$, the design variables are c and h, their bound are $c \in [0.1, 0.5]$ m and $h \in [0.1, 0.5]$ m according to the practical requirements. Then, we can obtain K_{global} stiffness distribution as shown in Fig. 8.8. From Fig. 8.8, one can see that the maximum K_{global} is around 4100 N/m. Similarly, when we assume $r = 0.2$ m, $\varphi = 20°$, $\psi = 25°$, $\eta_i = 1$, the global stiffness K_{global} distribution is shown in Fig. 8.9. From the figure one can see that the maximum K_{global} is around 4000 N/m. When the design variables are φ and ψ, and assuming different values of r, c, and h, one can also have the K_{global} distribution. Through different experiments we found that r, φ, ψ, c, and h dramatically affect the K_{global}. Next, we need to determine the values of the c, h, r, φ, and ψ simultaneously that lead to the maximum K_{global}.

3.4.2 Optimization

Genetic algorithm (GA) is based on the natural selection, and it repeatedly modified a population of individual solutions. At each step, the GA selects individuals at random from the current population to be parents and use them produce the children for the next generations. Genetic algorithms have the advantages of good convergence and robustness properties [8].

The genetic algorithm is employed here to optimize the global stiffness K_{global} of the bio-inspired hybrid mechanism. Our objective function is: $K_{\mathrm{global}} = \eta_1 k_{11} + \eta_2 k_{22} + \eta_3 k_{33}$, our purpose is to maximum the K_{global}. Note that the optimization functions in the GA minimize the objective function, in order to maximize the
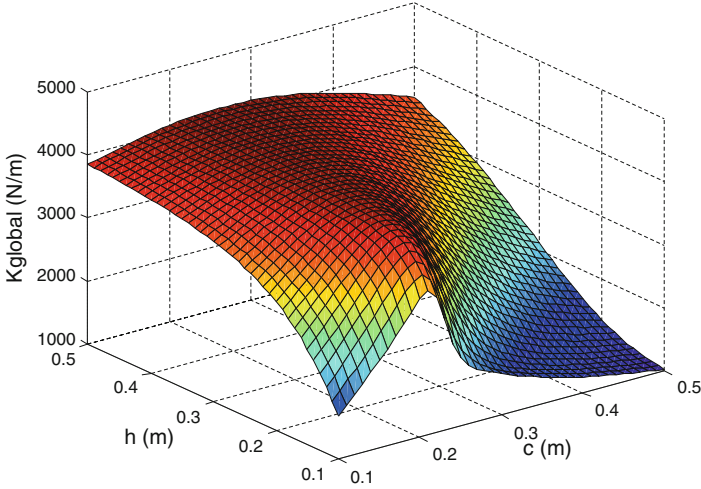
Fig. 8.9 K_{global} distribution when $r = 0.2$ m, $\varphi = 20°$, $\psi = 25°$

objective function, we need to minimize $-K_{global}$ because the point at which the minimum of $-K_{global}$ occurs is the same as the point at which the maximum of K_{global} occurs. The design variables are c, h, r, φ, and ψ. Their bound are the following according to practical requirements,

$c \in [0.1, 0.5]$ m, $h \in [0.1, 0.5]$ m,
$r \in [0.1, 0.5]$ m, $\varphi \in [-30°, 30°]$, $\psi \in [-30°, 30°]$

Some genetic parameters and operators are set as:

Scaling function: Proportional
Selection function: Roulette
Crossover function: intermediate
Crossover ratio: 1.0
Mutation function: adaptive feasible
Population size: 20
Maximum number of generations: 100

Run the optimization and we can have the following results:

The results of the global stiffness optimization are shown in Fig. 8.10. The points at the bottom of the plot represent the best fitness values, and the points above them are the mean fitness values in each generation. The plot also displays the vector entries of the individual with the best fitness function value in each generation. The optimal parameters are obtained after 51 generations as follows:

$$[c, h, r, \varphi, \psi] = [0.496 \text{ m}, 0.476 \text{ m}, 0.491 \text{ m}, 0.002 \text{ rad}, -0.037 \text{ rad}]$$

And the maximum K_{global} is 4983.1253 N/m
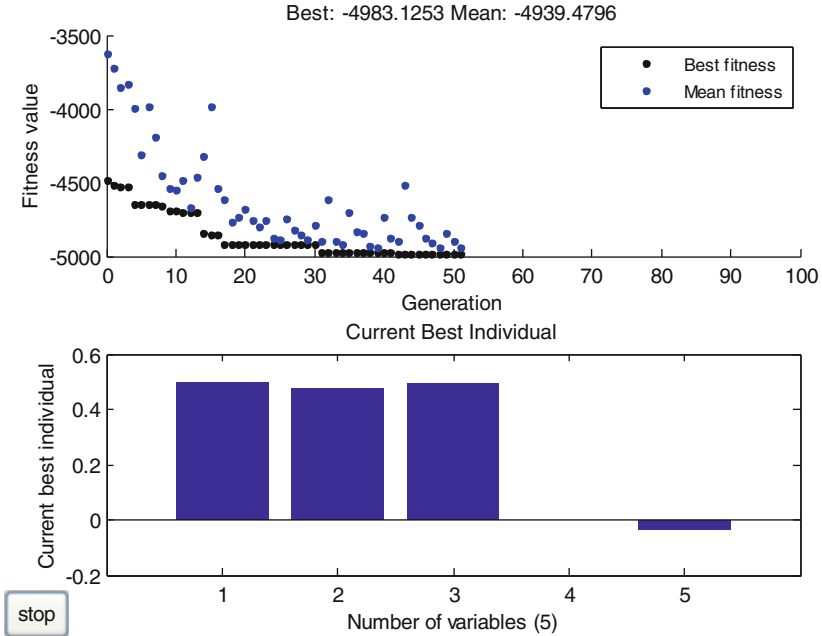
Fig. 8.10 The best fitness value and the best individuals of the global stiffness optimization

The results suggest that in order to make the global stiffness of the manipulator reach the maximum, the radii of the moving platform should be 0.496 m, the heave of the manipulator should be 0.476 m, the radii of the base should be 0.491 m, $\psi = 0.002$ rad, and $\psi = -0.037$ rad.

4   Gas Detection and Early Warning Technology Research

4.1   Preparation of Nano-Catalytic Element

By using two nano-catalyst preparation methods, i.e., precipitation impregnation and co-precipitation methods, two different nano-catalyst materials are prepared, I-Pd/Al$_2$O$_3$ and M-Pd/Al$_2$O$_3$. Two specific surface area of nano-catalytic materials are larger than ordinary catalysts, and the specific surface area of catalysts prepared by co-precipitation is better than that of precipitation dipping. From the initial temperature Ts and complete reaction temperature Tc, we know that the activity of the catalytic element containing nano-catalytic is better than that of ordinary elements, and co-precipitation method is slightly better than the precipitation dipping (Table 8.1).

Table 8.1 Specific surface area comparison between materials

Catalyst	Specific surface area 400 °C (m^2/g)	Specific surface area 650 °C (m^2/g)
C-Pd/Al$_2$O$_3$	98.2	52.3
I-Pd/Al$_2$O$_3$	216.4	175.4
M-Pd/Al$_2$O$_3$	228.7	168.5

Fig. 8.11 Catalytic sensor

4.2 The Effect of Doped Cocatalyst on Catalytic Component Performance

By using Sol-gel method [9], three kinds of carrier materials Ce-Zr-Al$_2$O$_3$, Ce-Al$_2$O$_3$, and Zr-Al$_2$O$_3$ are prepared, and by impregnating an equal volume of Pt and Pd solution to prepare catalytic elements. Experimental results show that, with the prepared carrier by Ce, Zr, and Al$_2$O$_3$, and due to the addition of Zr, the oxygen storage capacity of CeO$_2$ was greatly improved, and the advantages of a strong activity of CeO$_2$ was retained [10]. Because of the synergy effect of Ce-Zr, the phase transition from γ-Al$_2$O$_3$ to α-Al$_2$O$_3$ was hindered, the specific surface area was stabilized, and the sensitivity and stability of gas sensors was ensured (Fig. 8.11).

4.3 Effect on Catalytic Component Performance by Adding Pt-Pd

The specific surface area of the catalyst that contains metal Pd is slightly larger than that of catalysts Pt/Al$_2$O$_3$ that only contains Pt. In addition to [1]Pd$_2$Pt/Al$_2$O$_3$, the specific surface areas of other two Pd alloy catalyst are also larger than that of single metal catalyst Pd/Al$_2$O$_3$. The overall trend for the specific surface area is reduced while the active component Pt is increased. The catalyst of Pd: Pt for 2: 1 ratio has the highest catalyst-specific surface area. From the trend of dispersion, with the increase in the overall proportion of Pt in the catalyst, the degree of dispersion will increase, and monometallic Pt/Al$_2$O$_3$ has the highest degree of dispersion.

4.4   Cold Plasma Modification Process and Its Impact on the Performance of the Catalytic Element

After the glow plasma modification of Pt-Pd/Al_2O_3 material, helium is excited to high-energy cold plasma, and its ionic internal temperature can reach more than 104 K. The specific surface area is roughly 60 m^2/g larger than that of ordinary nanomaterials. The degree of dispersion of metal Pt-Pd on the carrier alumina also increased by about 31 %.

5   Conclusions

Mine safety exploring and rescuing robot developed in this chapter includes a system with certain function of drilling and excavation, a sensing system, and an external support system of wireless communication. The maximum and minimum stiffness distribution trend and the global stiffness optimization for the head section of type III rescue robot are investigated in details.

References

1. Tsai LW, Joshi S (2000) Kinematics and optimization of a spatial 3-UPU parallel manipulator. J Mech Des 122:439–446
2. Tsai LW, Joshi S (2002) Kinematic analysis of 3-DOF position mechanisms for use in hybrid kinematic machines. J Mech Des 124:245–253
3. Zhang D, Gosselin CM (2001) Kinetostatic modeling of N-DOF parallel mechanisms with a passive constraining leg and prismatic actuators. J Mech Des 123:375–381
4. Li YM, Xu QS (2008) Stiffness analysis for a 3-PUU parallel kinematic machine. Mech Mach Theory 43:186–200
5. Zhang Q, Zhang D, Cao H, et al (2011) Bionic peristaltic based walking mechanism and its control method. Patent CN 201110318141
6. Feng Y, Sun Y, Cao H, et al (2011) A mine rescue robot with driving function and its control method. Patent CN 201110370250
7. Zhang D, Bi ZM, Li BZ (2009) Design and kinetostatic analysis of a new parallel manipulator. Robot Comput Integr Manuf 25:782–791
8. Zhang D (2009) Parallel robotic machine tools. Springer, Boston
9. Wang Y, Tong M, Zhang T (2010) Improving anti-sulfur performance of methane sensors by new vectors and active components. Min Sci Technol 20(6):913–916
10. Wang Y, Tong M, Zhang D (2011) Improving the performance of catalytic combustion type methane gas sensors using nanostructure elements doped with rare earth co-catalysts. Sensors 11(1):19–31

Chapter 9
Reconfigurable Robot Manipulators: Adaptation, Control, and MEMS Applications

Gokhan Gungor, Baris Fidan, and William W. Melek

Abstract Reconfigurable robotic systems offer adaptable hardware which can modify its configuration based on the objectives to be accomplished. They consist of multiple individual modules. These modules, when combined, allow efficient and feasible reconfiguration. However, control of reconfigurable robotic manipulators is particularly challenging because of the inherent nonlinearity, unknown parameters, and uncertainties and variations in the system dynamics. The uncertainties and variations are mainly due to setup reconfigurations, friction and varying payloads, and are typically addressed using adaptive control. The purpose of this chapter is to present a systematic review of key control schemes for reconfigurable robotic systems, highlighting their benefits and disadvantages. We start with an introduction to the state of the art of reconfigurable robots and their applications. We also review the application of these systems at micro-scale. Then, we review control architectures for such systems, including linear, adaptive, and robust.

Keywords Reconfigurable robot manipulators • Micro-robot • Control

1 Introduction

Concise Robot Institute of America (RIA) defines a robot as: "A re-programmable, multifunctional manipulator designed to move material, parts, tools, or specialized devices through various programmed functions for the performance of a variety of tasks" [100]. A robotic manipulator is used for a variety of applications utilizing its job precision and rapid operations, and helping reduce the manufacturing costs in industrial applications for many years [64]. For example, the four-axis selective compliant assembly robot arm (SCARA) is used for the precision in the assembly line to reduce the production cost and to save time in a wide variety of industrial areas. Moreover, the number of degrees of freedom (DOF) can be increased to extend the flexibility of robot manipulators. For instance, the programmable

G. Gungor (✉) • B. Fidan • W.W. Melek
University of Waterloo, 200 University Ave W, Waterloo, ON, Canada N2L 3G1
e-mail: ggungor@uwaterloo.ca; fidan@uwaterloo.ca; william.melek@uwaterloo.ca

universal manipulation arm (PUMA) is mostly used for quality improvement of welding applications in a broad range of manufacturing and automotive areas [4, 92]. Robot manipulators are re-programmable systems in nature. However, because of their fixed structures, the operation envelope and reach of conventional robot manipulators are restricted [103]. Reconfigurable robot manipulators (RRMs) increase flexibility, dexterity, and ability to complete tasks that cannot be performed by such fixed configuration manipulators [83].

The term "small-scale robotics" is used for the non-conventional robotic systems that range from millimeters to nanometers in diameter [96]. In the literature, there exist important small-scale robotic applications in medicine and manufacturing areas. However, such robots at this level involve many challenging aspects in their fabrication, control, and power transmission [96]. In recent years, a primary objective in the reconfigurable robotics research focused on minimizing the size of the robotic modules, actuators, and power supplies. Reduction in the size of modules brings new challenges such as module manufacturing, control, and communication [99]. Moreover, in electric actuators, the torque required to rotate an object increases dramatically when the link length decreases, causing loss of efficiency [104]. Small actuators have limited use for the micro-scale robots because of their low power/weight ratio [96, 147]. Hence, these drawbacks necessitate the development of microelectromechanical systems (MEMS) technology to fabricate tiny actuators, sensors, and mechanisms that can be used to build RRMs at this scale of operation.

RRMs require a proper control design to achieve the precision and stability required for tasks such as point-to-point and path planning because their system dynamics mostly involve nonlinearities and uncertainties due to reconfigurability, friction, and varying payloads. Unknown physical parameters, unpredictable behaviors of the system in the environment, and inaccurate fabrication of robot components cause further control problems for micro-robot applications that are under 1 mm in size. Most of the RRM control schemes in the literature have been designed at macro-scale, aiming to compensate position errors and achieve accurate motion during the task. Such control designs include linear, adaptive, and robust ones.

The focus of this chapter is to review the control schemes developed in the literature for macro and micro-scale RRMs. The chapter is organized as follows: Sect. 2 introduces to the state-of-the-art RRMs and their applications, focusing on micro-scale robots. Section 3 reviews existing control architectures for such systems. Section 4 presents control applications for small-scale robots. Conclusions are presented in Sect. 5.

2 Reconfigurable Robot Manipulators

Conventional robotic systems have operational limitations in many applications because their fixed structures are not convenient for conversion of joint axes to accommodate various tasks. These limitations have increased the demand for RRMs in many industries such as automotive and healthcare. RRMs are usually

made up of similar components of conventional robots such as actuators, sensors, power supplies, communication devices, and processing units. In addition to all of these, they also have functional components such as active/passive connectors for connection or disconnection of modules [2].

Moreover, they consist of independent and typically identical modules. Each module is capable of independent movement and can be connected to other modules to allow formation of various required configurations [19, 23, 99]. Modification of kinematic and dynamic parameters of the system is essential to execute desired configurations to perform specific tasks. In this section, we present a summary of RRMs along with their concepts of operations. Several examples of the RRMs developed in the literature are shown in Fig. 9.1. In an RRM, the individual modules can be added and removed to generate alternate configurations [12]. There exist three categories of modules for RRMs: actuator modules, link modules, and joint modules [21, 135]. Joint modules are critical components of such robots. Common joint modules that are used to construct RRMs are rotary, pivot, and spherical. Various link shapes and sizes can be employed to generate different configurations
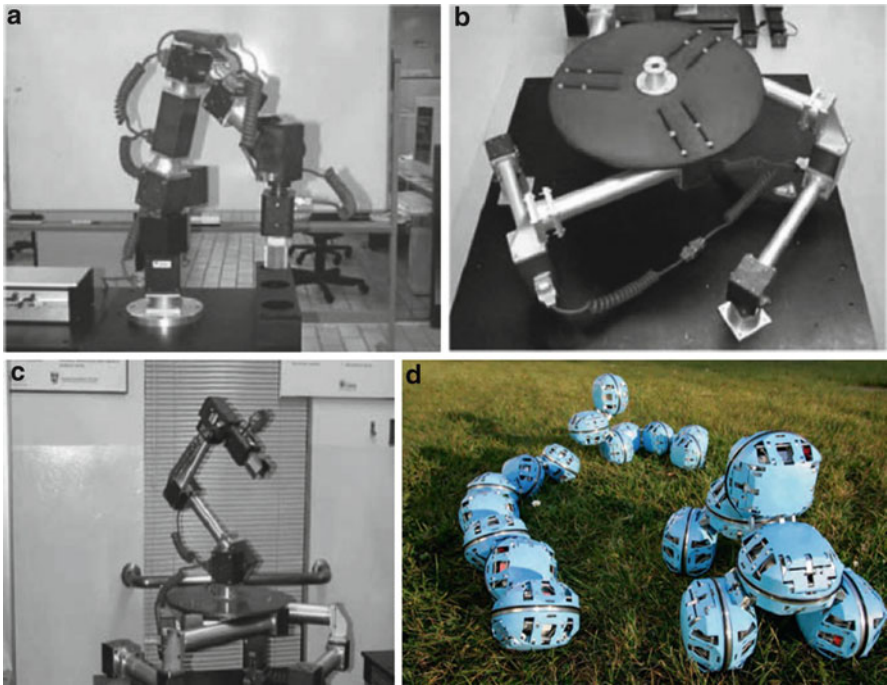

Fig. 9.1 Some examples for reconfigurable robot manipulators in the literature. (**a**) A reconfigurable serial robot [30]. (**b**) A reconfigurable parallel robot [30]. (**c**) A hybrid (combination of serial and parallel robot) reconfigurable robot [30]. (**d**) An example of self-reconfigurable robot which has one type of module [61]

by reassembling them for various specific tasks. Detailed properties of links and joint-actuator modules can be found in [21, 135].

RRMs can perform a great number of jobs uncertain or unstructured environments that range from precision machining to nuclear industry [12, 97, 152]. This is the main advantage of RRMs; their adaptability. Besides, the other benefits of RRMs are low cost, simple maintenance, transportability, and durability [82].

Since the introduction of the concept of modular robots in the early 1990s, various models and types have been introduced and improved overtime. Modular RRMs (MRRMs) generally are composed of three groups [12] which are: modular reconfigurable serial robots (MRSRs) [3, 11, 21, 82, 103], modular reconfigurable parallel robots (MRPRs) [22, 24, 30, 38, 135–138, 149], and modular self-reconfigurable robots (MSRRs) [15, 36, 50, 60, 87, 153].

2.1 *Modular Reconfigurable Serial Robots*

MRSRs are a type of serial-link manipulators. These are capable of performing a lot of serial configurations. They can alter their kinematics, dynamics, and control properties to form numerous tasks in comparison with their conventional counterparts by adding or removing joints, arranging their positioning, and adjusting the control gains. The general building blocks of an MRSR are link modules, joint modules, and control unit. For example, the TOshiba Modular Manipulator System (TOMMS) [82] is composed of several joint modules of the same type, several link modules of the same type, and a control unit. Despite the use of identical joint modules and identical link modules, several robot configurations can be obtained by only rearranging the joints and links according to the task requirements. A schematic representation of MRSRs [82] is depicted in Fig. 9.2.

Fig. 9.2 Schematic presentation of the TOMMS with joint and link modules in [82]. There is also a depiction of two possible configuration of the TOMMS system in this figure
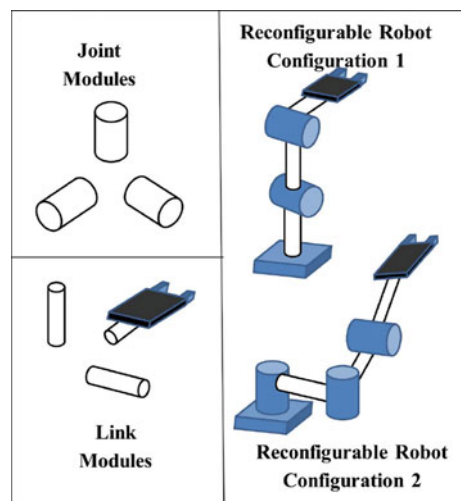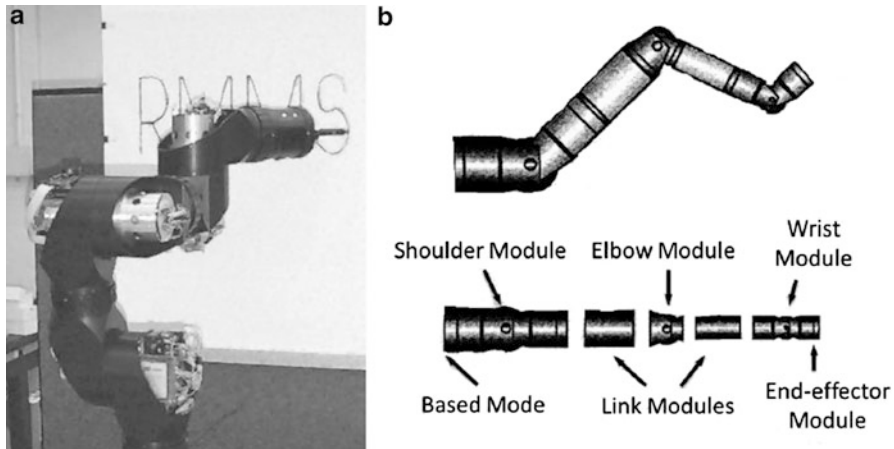
Fig. 9.3 Two examples for MRSR. (**a**) An example configuration for reconfigurable modular manipulator system (RMMS) robot [97]. (**b**) 7-DOF modular serial manipulator [16]

In the literature, there is a great deal of work on MRSR such as [3, 16, 26, 27, 54, 70, 82, 97, 154] (Fig. 9.3). One of the MRSRs developed in [103] comprises self-contained modules and a control software. Li et al. [70] present an MRSR with the joint module that can perform either rotational or pivotal motion with four connection ports. In the following subsection, we review MRPRs available in the literature.

2.2   Modular Reconfigurable Parallel Robots

Parallel robotic platforms can be used for many applications such as precision machining, precision assembly, and surgical operation. Such parallel platforms have several advantages in comparison with their serial counterparts, including high payload carrying ability, velocity accuracy, stiffness, and lower inertia. However, they have some drawbacks such as small workspace and singularities in its workspace [43, 84]. Details of kinematics and dynamic properties of parallel robots can be found in [84, 150].

MRPRs comprise independently modeled modules that can be formed into several configurations [30]. Benefits of the MRPR platforms such as ease of maintenance, flexibility, and transportability have been pointed out. A significant number of different MRPRs exist [17, 22, 24, 30, 38, 66, 93, 137–140, 149, 156]. Dash et al. [30] propose two different examples of MRPRs shown in Fig. 9.4. In addition, a robotic work cell [20] is constructed with MRSR and MRPR platforms. These platforms have conventional and functional components. Besides, this robotic work cell is an attractive example for many industrial applications because it brings serial and parallel robotic systems together within a single work cell.
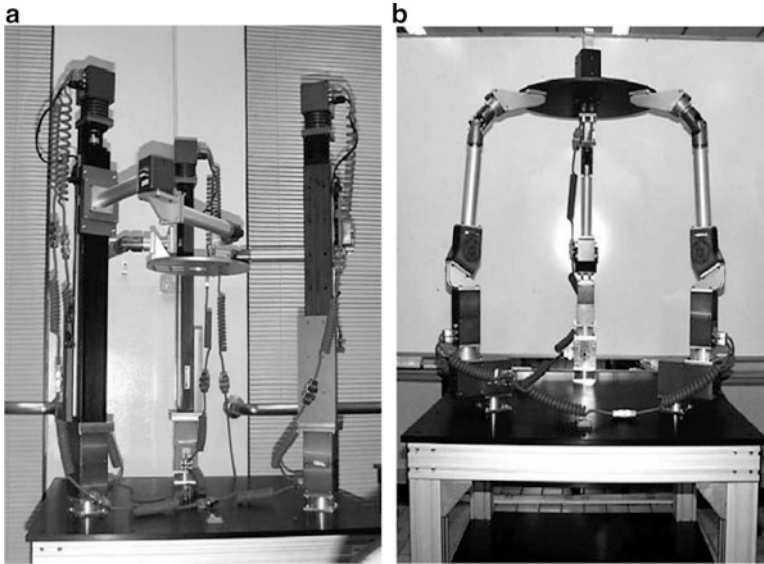
Fig. 9.4 Two different examples reconfigurable parallel manipulators: six-DOF and three-legged [30]. (**a**) RPRS three-legged reconfigurable parallel robot. (**b**) RRRS three-legged reconfigurable parallel robot

Reconfigurable hybrid parallel robots have also been proposed and studied in the literature using features of MRSR and MRPR. Coppola et al. [28] present a six degree of freedom and Palpacelli et al. introduce the design of a reconfigurable 3-DOF parallel kinematics manipulator, whose primary feature is changing its mobile platform from translation to rotation [93].

2.3 Modular Self-Reconfigurable Robots

MSRRs usually consist of a large number of identical active modules that can rearrange themselves in a large variety of configurations or structures. According to Whitesides and Grzybowski [133], self-assembly robotic systems are reconfigured automatically into various configurations at runtime because their modules have special features to attach or detach themselves without human inference. MSRRs offer three essential features: versatility, robustness, and low-cost [144]. They are superior to fixed configuration robots particularly in uncertain/unstructured environments, because their modules have abilities such as being programmable to act as agents. There exist several different self-reconfigurable modular systems in the literature to generate 2D [88, 129] and 3D [87–130] geometrical structures and motions.

In the literature, there exist several well-known MSRR systems [25, 40, 47, 61, 94, 101, 102, 121, 128, 132, 142, 146]. A cellular/modular robotic system (CEBOT)

a

b

A module representation of MICHE
robot [60]

An example configuration with MICHE
modules [41]
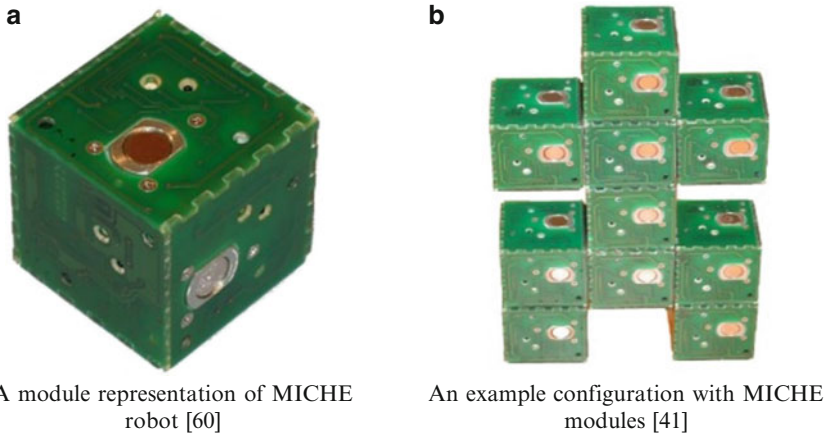
Fig. 9.5 (**a**) A 1.8-inch autonomous cube-shaped (**b**) a complex configuration of MSRR

a

b

An example configration with
MTRAN 2 modules [123]

An example configration with
MTRAN 3 modules [122]
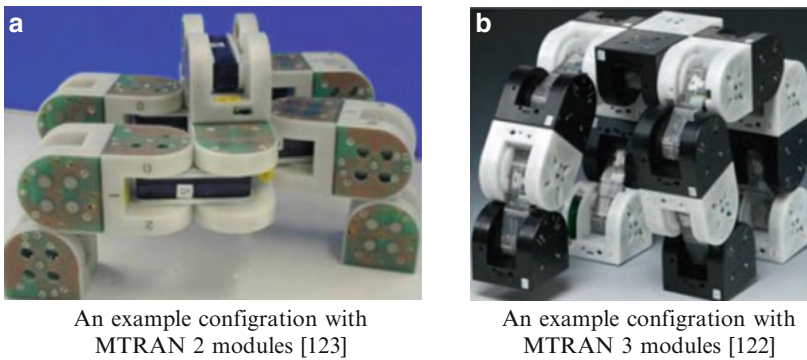
Fig. 9.6 Complex shape (**b**) with a 1.8-inch autonomous cube-shaped (**a**) [47]

is proposed by the researchers in [41, 42]. Gilping et al. proposed the autonomous
module shown in Fig. 9.5a. A complex shape with identical modules in Fig. 9.5b. has
achieved in the study [47]. Murata et al. [87] have proposed a modular transformer
self-reconfigurable robotic system (M-TRAN). M-TRAN as shown in Fig. 9.6 was
designed with identical robotic modules to form a variety of configurations. The
paper [143] has introduced Connector Kinetic roBot (CKbot), shown in Fig. 9.7.
This robot has the ability to recover itself after being partitioned into many pieces.

2.4 Micro-Scale Reconfigurable Robots

In recent years, development of MEMS technologies allowed researchers to scale
down robotic systems. Miniaturizing the actuation system, robot mechanism, and

Fig. 9.7 Two different configurations with CKbot modules [47]

power sources are essential for the advancement of a micro-scale robotic system. However, micro-robotic systems are generally driven and sourced using external actuation and power systems. These robots are used in numerous applications such as health care/micro-medicine [5, 18, 39, 56, 124, 141], and micro-optics [7, 48]. According to [1], micro-robotic system research areas can be classified into two categories: micromanipulation systems (Fig. 9.8) and microfabrication/micro design.

To carry out required tasks, micromanipulation systems [10, 44] handle objects that range from micrometer to millimeter in diameter [1]. Micro-scale manipulator systems are different from the macro-level ones in terms of the scaling of physical effects. This issue poses a challenge for the micro-scale systems. For example, the effect of the surface forces is more dominant at the micro-scale because surface/volume ratio increases when the size of the object decreases. Therefore, initial forces and mass have less influence on motion dynamics whereas heat transfer, fluid dynamics, surface chemistry, and adhesion based forces become significant on micro-scale robots. This issue poses a challenge in position control for micromanipulation systems because physical laws used for macromanipulation are not applicable to micro level ones [1].

Miniaturization of actuators for motion and power generation for the actuators are important issues for such systems [1]. At this level, the off-board or external actuation is needed since integration of a power source and actuation system into micro-scale modules is current difficulty in the field [114, 115].

There exist many methods in the literature to externally actuate the micro-robots. The control methods for planar motion are based on laser thermal excitation [122], electromagnetics [98, 131], and electrostatics [34]. For example, the design in [99] controls the multiple modules using an electrostatic anchoring surface. There are further methods to actuate the micro-robots in three-dimensions electromagnetically [8, 37, 80]. A swimming micro-robot called example, OctoMag, actuated by a magnetic gradient with wireless connection can be seen in [67]. The magnetic system for OctoMag contains eight soft-magnetic-core electromagnets to produce the required magnetic force. There exist some helical micro-robot designs actuated with rotating magnetic field in the literature. In [46, 151], rotational motions of the robot are converted to linear displacements. Microrobotic assembly can be

Fig. 9.8 Micro-robot manipulation with (**a**) gripping, (**b**) impulsive forces, (**c**) vacuum forces, (**d**) surface tension, (**e**) electrostatic forces [1]
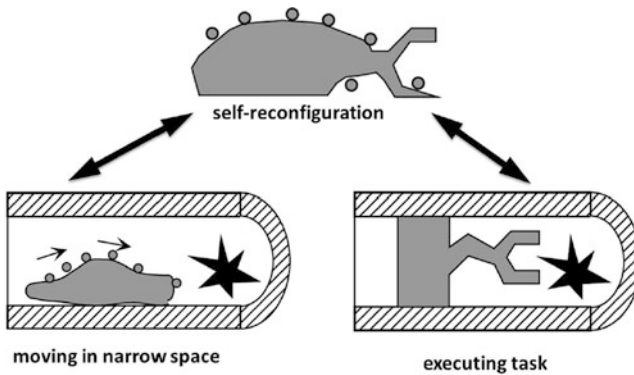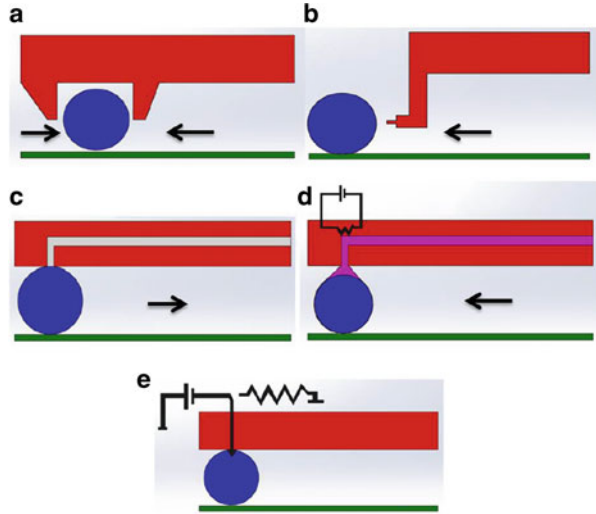
Fig. 9.9 Applications of a micro reconfigurable robotic system [146]

performed using several microrobots that are actuated by electromagnetic [34]. However, such micro-robots do not have ability to reconfigure themselves. All aforementioned actuation methods have some challenges. For instance, piezo-electric actuation requires high voltages and high power actuation.

The design in [145] contains one of the smallest modules (2 cm diameter) constructed with built-in actuator. This module in Fig. 9.10 is driven by shape memory alloy (SMA) actuator. Mobes et al. [86] presents a miniaturized 2D modular and self-reconfigurable robot for conveying the micro parts (Smart Blocks) that are smaller than 1 cm^3 in volume. In the literature, there exist many examples for micro self-reconfigurable robots (Fig. 9.9) [31, 33, 49, 51]. For instance, Green et al. [49] and Hosokawa et al. [51] propose micro-scale self-reconfigurable systems for

Fig. 9.10 Structure and prototype of micro module for reconfigurable robot [145]
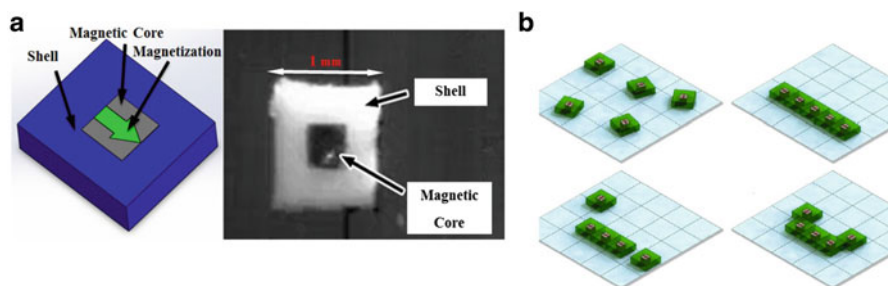
Fig. 9.11 (**a**) Photograph of a Mag-μMod Complex shape (**b**) Five Mag-μMods operating on an electrostatic grid surface [31, 99]

micromanufacturing. However, these systems are not be controllable with surface driving force.

In the last few years, micro level modules under 1 mm in size have been fabricated utilizing advanced MEMS technology for reconfigurabl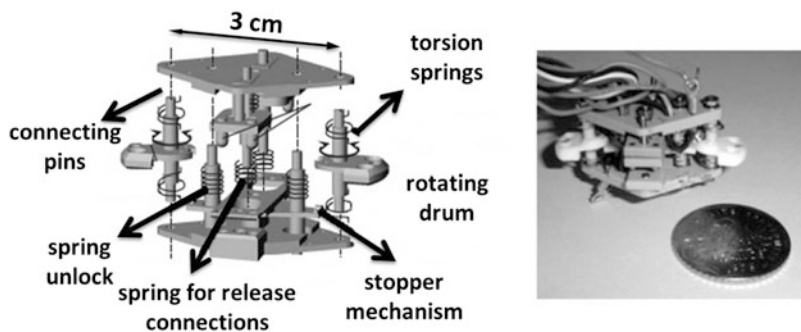e robot applications. The studies in [31, 99] present reconfigurable micro-scale robots design Mag-μMods composed of permanent magnet modules whose dimensions are $900 \times 900 \times 270\,\mu$m as shown in Fig. 9.11a. These modules are capable of moving on a planar surface. Besides, the five magnetic micro modules are assembled and disassembled using electrostatic and electromagnetic fields as shown in Fig 9.11b.

According to [33], the significant interest in such robotic systems is based on the ability to create arbitrary 2D/3D shapes, mechanical strength, electrical conductivity/continuity, reconfiguration, and disassembly (Fig. 9.12). The authors in [33] used permanent magnet modules with diameters less than 1 mm to move a reconfigurable micro-robot on 2-D surfaces using oscillating external magnetic fields of several mT in strength. Diller et al. [33] have proposed a magnetic bonding method which offers increased mechanical strength for this robot structure. This method allows out-of-plane 3-D physical motion and electrical conductivity between modules.

Fig. 9.12 Micro-scale modular robot concept [145]

3 Adaptive Control of Reconfigurable Robot Manipulators

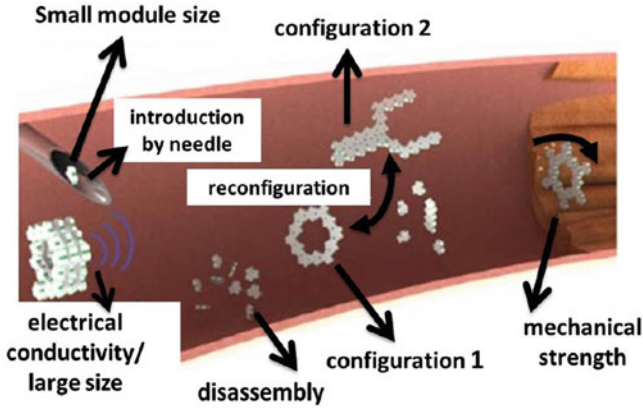Two key goals in control design for RRMs are precision and stability. Control of RRMs is particularly compelling for researchers in the field because of nonlinearities, unknown parameters, and uncertainties in their systems. There has been a considerable interest in designing efficient control schemes for such systems. The purpose of this section is to review the literature on control of MRSR platforms. An example of reconfigurable robot control structure is shown in Fig. 9.13.

3.1 A Discussion on Adaptive Control of Robot Manipulators

Adaptive control is an approach to tune parameters (on-line) of the nominal controller of a dynamical system, to compensate effects of nonlinearities, uncertainties, and variations in the system parameters. The survey in [6] shows that adaptive control design with sufficient practical precautions is used successfully in many applications. In [58], the adaptive control methods are grouped into two broad types: identifier-based and non-identifier-based schemes. The identifier-based scheme is classified into two different approaches as direct and indirect. Indirect adaptive approach is based on estimation of unknown or uncertain plant parameters; and the plant parameter estimates are used for the calculation of the control parameters as shown in Fig. 9.14.

Indirect adaptive control, the closed loop system is parametrized in terms of the control parameters to be tuned; and these parameters are tuned adaptive, directly considering the control task. Direct adaptive control structure is depicted in Fig. 9.15. In addition to these two structures, there also exist non-identifier-based adaptive control methods. Gain scheduling is one of such control methods:
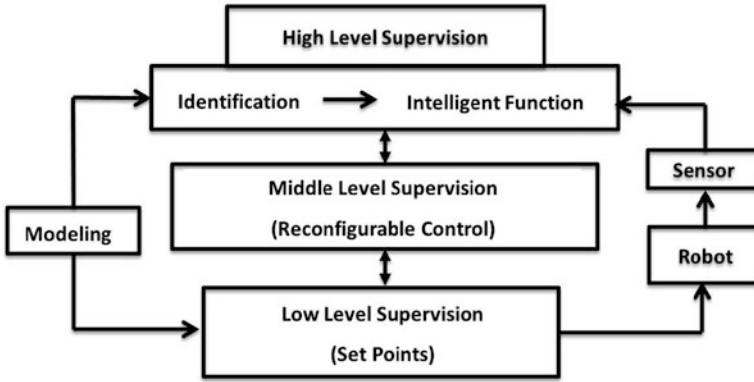
Fig. 9.13 Reconfigurable robot control structure [77]
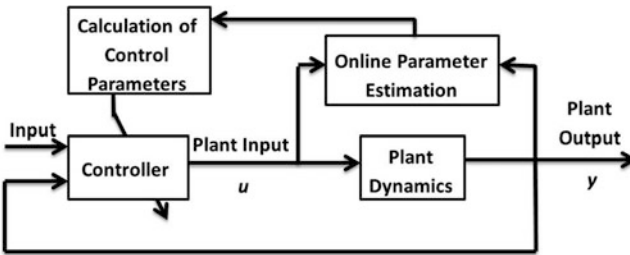
Fig. 9.14 Indirect adaptive control structure
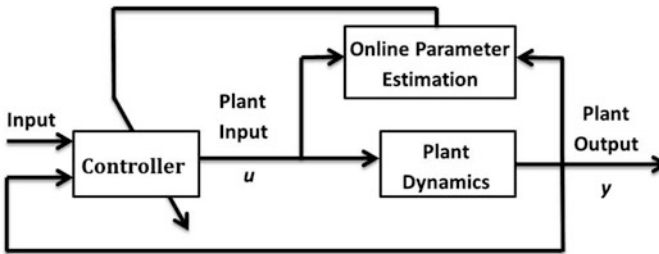
Fig. 9.15 Direct adaptive control structure

It consists of a lookup table, gain scheduler logic, and selecting a proper constant feedback controller gain from lookup table [58].

For robot manipulators, various direct, indirect, and composite adaptive controllers have been designed in the literature, see, e.g., [69, 85, 117]. Slotine and Li [117] present a direct adaptive control scheme to eliminate the error of each joint motion with an adaptive law. In [69], the main aim of using adaptive law is reducing the tracking error of the system. In [85, 117], the authors present an

adaptive feedforward-plus-PD controller with the direct method. The controller design in [85, 117] does not require acceleration measurement and some excessive calculations.

Indirect adaptive controller approach [69] is utilized to obtain the nominal values of uncertain/unknown parameters using prediction errors on the joint torques. Li and Slotine [69] proposed four different parameter estimation methods that are the gradient, least square, the gain-adjusted-forgetting estimator, and inherently-bounded-gain estimators for the adaptation mechanism. These are used as adaptation mechanisms to provide convergence of the controller parameters to their nominal values in adaptive control. Middleton et al. [85] present a computed torque control design and a least squares estimator based on an input–output stability analysis to track the required trajectory. Hsu et al. [53] design an indirect adaptive controller, composed of a gradient parameter estimator algorithm and a computed torque control. Slotine [116] explores a robust control scheme to decrease tracking errors in presence of uncertainty and disturbance in the robot manipulator dynamic model. Adaptive computed torque method is integrated with an on-line estimation algorithm to eliminate nonlinear effects in the robot dynamic in [29]. Hun et al. [55] have proposed an adaptive backstepping controller for robot manipulators in the presence of uncertainty and external disturbance. Many studies exist to achieve efficient trajectory and the robustness for robot manipulator [118, 119]. Decentralized controller design is also a popular method for many robotic systems according to most of the researchers in the field due to its computation simplicity [52]. Adaptive decentralized control approaches have been designed for linear and linear dominant systems in the literature as well [45, 57, 110].

3.2 Adaptive Control of Macro-Level Serial Type Reconfigurable Robots

The dynamic equation given in Eq. (9.1) for an n-DOF serial robot is derived from Euler–Lagrange formulation [155]. It is also applicable to any configuration in RRM but $M(q)$, $C(q, \dot{q})$, and $G(q)$ parameters will be different in each configuration.

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) + F(\dot{q}) = u \tag{9.1}$$

where $q = [q_1, q_2, \ldots, q_n]^T \in \Re^n$ is defined as joint position; $\dot{q} = [\dot{q}_1, \dot{q}_2, \ldots, \dot{q}_n]^T \in \Re^n$ is the joint velocities; $M(q) \in \Re^{n \times n}$ describes inertia matrix; $C(q, \dot{q})\dot{q}$ is centripetal and Coriolis force; $G(q) \in \Re^n$ presents the gravity part; $F(\dot{q}) \in \Re^n$ is the frictional force; and $u \in \Re^n$ is the input control. The desired position trajectory is $q_{ei}(t) \in \Re$ and the tracking error is $e_i = q_i(t) - q_{ei}(t)$.

The decentralized control method also called independent joint control since each joint is considered as an independent subsystem. It is more appropriate than other control methods for RRMs. In this method, dynamic effects from the other links

and joints are treated as disturbances. Moreover, decentralized control has some advantages such as robustness against failures, computational efficiency, and design simplicity [111]. It enables exchange of information among subsystems to design the controller with only local measurement variables [125]. However, applying decentralized methods are not simple for specific tasks as stated in [52] because the control inputs and states are decoupled from each other due to presence of the dynamic couplings in the system dynamics.

Implementation of the decentralized control method into reconfigurable robot systems has become popular [35, 75]. One of the related control designs is proposed by Liu et al. [74] in which a modular distributed control for an MRSR based on joint torque sensing is presented. It is claimed that the control parameters of the modules do not require an adjustment when the modules are reassembled for the various specific tasks. Moreover, the model uncertainties such as payload, joint friction, and unmodeled dynamic couplings are resolved using joint torque sensor measurements and robust controller design.

Another study conducted by Zhu and Li [155], a decentralized adaptive fuzzy sliding mode control scheme is designed. In the study, the Takagi–Suego fuzzy logic system with an adaptive algorithm is proposed to approximate unmodeled dynamics of MRSRs. Additionally, an adaptive sliding mode controller is designed to minimize fuzzy approximation error and handle the uncertainties in the system interconnections.

Li et al. [72] present a decentralized robust controller for MRSRs based on the Lyapunov stability analysis and backstepping technique. Since the defined controller is configuration-independent, it enables fast reconfigurability of the manipulator and a control strategy to achieve precise position tracking during the required task. However, the need for joint modeling is the main drawback of the control design because derivation of joint dynamics is challenging due to existence of nonlinearities and uncertainties.

In addition to aforementioned control approaches, neurofuzzy intelligent control design is presented in [83] for MRSRs. The authors use learning control with on-line feedforward neural-network to compensate for unmodeled system dynamics due to alternating robot configurations. A skill module updating the adaptive control parameters is also included in the controller design. To achieve accurate position tracking of an MRSR with unknown dynamic parameters, internal Type-2 Takagi–Sugeno–Kang fuzzy logic and PD controllers are used in [12] to handle uncertainties, unknown parameters, and dynamics of the system.

Joint velocity and torques can be measured by sensors such as tachometer and strain gauge, respectively. However, the inclusion of additional velocity and torque sensors to the system would increase the cost and system complexity. In addition, such devices mostly create noise problems that could not be filtered away easily. On the other hand, it is possible to estimate joint torque and velocity for MRSRs using observer design [73]. An example of such approach is given in [152], which presents a decentralized control scheme for RRMs without velocity sensor. The precise position signal in the study is utilized to reconstruct the velocity signal for each module joint. Moreover, the controller with adaptive fuzzy logic is used to

approximate the model uncertainties. Likewise, Li et al. [71] designed a fuzzy radial basis function (RBF) neural network compensation algorithm based on a velocity observation model to tackle the external disturbances and unknown dynamics.

4   Micro-Scale Robot Control

Micro-scale robots require a suitable control design to achieve accuracy and stability for a desired task. Micro-robots typically have unknown physical parameters, nonstructural behaviors in the environment, and imprecisely manufactured their components. These circumstances generate control challenges for micro-scale robots. For example, the geometric shapes and sizes of the robot components might not match the required robot properties because of inaccurate manufacturing in μm diameter. Hence, the uncertainty and inaccuracy of the physical parameters and unknown behaviors of the robot in the environment cause position errors for such system. These errors lead the robot to instabilities.

Efficient and robust control designs are significant for micro-scale robots because of presence of nonlinearities that come from predominant adhesion force, hysteresis phenomena, and environmental effects such as humidity noises. Adaptive control and neural networks or learning algorithms have been known as an effective approach for such system. Such controllers and estimation methods are put in the process when the system parameters and workplace have uncertainties [9, 14, 79]. In recent years, micro force sensing and control techniques were studied and developed [13]. Various motion control methods have been proposed [14, 46, 59, 65, 67, 78, 134, 151]. For example, the study in [105] presents an adaptive neural networks setup in Fig. 9.16 for manipulation of micro-cubes to achieve more accurate motion. The controller design is presented in [105] with the stability analysis based on Lyapunov method. Moreover, CCD camera is installed to measure the actual position of the system in the closed loop as shown in Fig. 9.17 [105–107].

The force magnitudes generally range from 10^{-3} to 10^{-6} N in micro-scale. Thus, it should be pointed out that force sensor is crucial to analyze the manipulation force in micro-scale. To satisfy this sensitivity, forces in the micro-level are measured by the sensing devices such as strain gauges, piezoelectric, and using laser-based optical techniques. In addition, there exist various detection devices in the literature such as laser scanning and cameras [14].

Micro-robots can be actuated by externally generated magnetic-field gradients to exert forces on the microsystems and drive them to the required positions. However, precisely estimating required forces is a significant challenge for many robot operations such as retinal surgery. In addition, the force exerted on a magnetic device depend on the position and orientation of micro-robot in the field and estimation errors in the system lead the micro device to instability [9, 79]. The position of the microsystem can be obtained using visualization system such as an optical microscope. There exist two different microscopy technologies in the literature: scanning electron microscope (SEM) [63] and atomic force microscope

Fig. 9.16 Schematic presentation of the experimental setup for pushing a micro size cube [105]

Fig. 9.17 Schematic presentation of the nonlinear adaptive neural network (NN)-based controller [105]

(AFM) [112] in Fig. 9.18. One of the most important features of both microscopes is having a presence of higher resolutions to watch the operation of the robot. Other main advantageous features of optical microscopes are on-line imaging, low cost, and adaptable workspace. However, they contain small workspace and some challenges for micro-manipulation or assembly process in 3D [63, 112].

Stochastic self-assembly approaches at the micrometer scale are reviewed in [81]. The assembly to execute different configurations and disassembly processes are stochastic in nature. The assembly process of such systems mostly require random excitation and a long periods of time. An example of micron-scale assembly using micro-scale robots is shown in [34]. In [34], the authors propose the assembly of four MEMS-fabricated silicon micro-robots. However, they cannot disassemble since the electrostatic fields do not allow for being partitioned into many pieces.

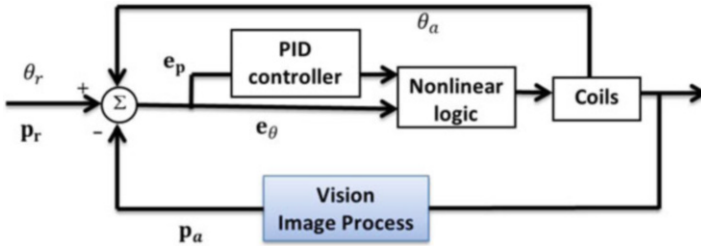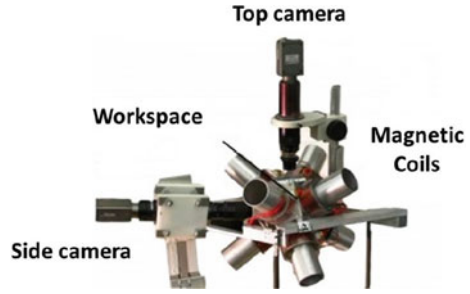**Fig. 9.18** Atomic force microscope from [99]

Fig. 9.19 Schematic presentation for control design of Mag-μMods [14, 32, 99]

The authors in [32] designed pushing based manipulation system using Mag-μMods as illustrated in Fig. 9.11. The Mag-μMod system allows for controlling the multiple micro-level robots. Here, the controller is designed using image processing that provides feedback information about the robot position to the controller at each instance. The designed controller consists of a combination of proportional-integrator-derivative (PID) and nonlinear logic as illustrated in the block diagram in Fig. 9.19. The main purpose of the nonlinear logic block is to adjust the orientation of the magnetic field θ and the velocity of the system. This is a requirement to satisfy the stability of the required motion. In this study, the robotic system is controlled in 2D motion [14, 32, 99]. In Fig. 9.19, error of orientation and position of the robot are described as e_θ and $\mathbf{e}_p = [e_{px}, e_{py}]^T = [p_{rx} - p_{ax}, p_{ry} - p_{ay}]^T$, $\mathbf{p}_r = [p_{rx}, p_{ry}]^T$ is the desired or required position and $\mathbf{p}_a = [p_{ax}, p_{ay}]^T$ is the actual position matrix that is obtained mostly from cameras or several sensors. This study assumes that the orientation of the Mag-μBot is similar to the orientation of the magnetic field.

OctoMag system is presented in [67] to create 3D movement. However, the system control is only provided for 5 DOF [14]. Magnetic field $\mathbf{F}_{\text{Bem}}(\mathbf{p})$ in the system is generated with electromagnets for each point $\mathbf{p}$ of the area. Unit current vector and scaler current value can be defined as $\tilde{\mathbf{F}}_{\text{Bem}}$ and I_{em}, respectively. An electromagnetic field is generated with an electromagnet, calculated as follows:

$$\mathbf{F}_{\text{Bem}}(\mathbf{p}) = \tilde{\mathbf{F}}_{\text{Bem}}(\mathbf{p})I_{\text{em}}. \qquad (9.2)$$
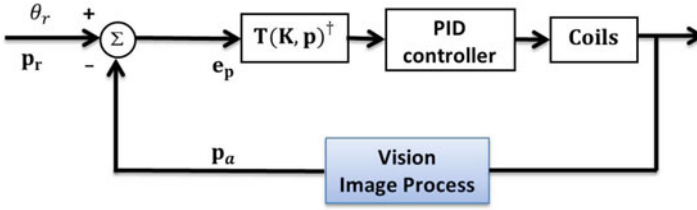
Fig. 9.20 Block presentation for control design of OctoMag [14, 67]

Moreover, the total electromagnetic field $\mathbf{F}_B(\mathbf{p})$ can be derived with a collection of each electromagnetic field as

$$\mathbf{F}_B(\mathbf{p}) = \sum_{em=1}^{k} \tilde{\mathbf{F}}_{\text{Bem}}(\mathbf{p})I_{\text{em}} = F_B(\mathbf{p})\mathbf{I}. \tag{9.3}$$

The magnetic field and magnetic force are formed in matrix form as follows:

$$\begin{bmatrix} \mathbf{F}_B \\ f_m \end{bmatrix} = \begin{bmatrix} F_B(\mathbf{p}) \\ \mathbf{K}^T\mathbf{F}_{Bx}(\mathbf{p}) \\ \mathbf{K}^T\mathbf{F}_{By}(\mathbf{p}) \\ \mathbf{K}^T\mathbf{F}_{Bz}(\mathbf{p}) \end{bmatrix} \begin{bmatrix} I_1 \\ \vdots \\ I_k \end{bmatrix} = \mathbf{T}(\mathbf{K}, \mathbf{p})\mathbf{I} \tag{9.4}$$

where $\mathbf{I}$ is a vector of the unit current $\mathbf{I}_{\text{em}}$, $\mathbf{K}$ can be defined as uniform magnetization matrix. Therefore, current $\mathbf{I}$ can be derived from $\mathbf{T}$, the force, and magnetic field of coils in the system as follows:

$$\mathbf{I} = \mathbf{T}(\mathbf{K}, \mathbf{p})^{\dagger} \begin{bmatrix} \mathbf{F}_{\text{Bref}} \\ f_{\text{mref}} \end{bmatrix}. \tag{9.5}$$

In [67], the position of the micro-robot is controlled with a closed loop system as shown in Fig. 9.20. In addition, it should be pointed out that open loop system is used to control the orientation of OctoMag robot since the micro-robot orientation will be the same as the orientation of the magnetic field when it is assumed that the orientation of the latter is not changing rapidly [14, 67].

Figure 9.21 demonstrates a 3D cubic shaped trajectory performed in an automated mode for the OctoMag robot. According to the result in [67], the OctoMag robot has standard deviations of 6.313, 4.757, and 8.951 μm along the x-, y-, and z-axes, respectively. To evaluate the trajectory tracking performance of the designed system, actual positions are observed with camera images based on 400 frames collected at 30 Hz.
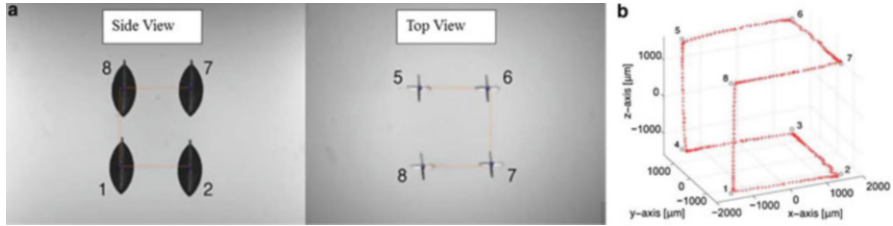
Fig. 9.21 (**a**) The side and top view of the microrobot during cubic shaped trajectory (**b**) tracking for cubic shaped trajectory in isometric graph with automated mode [67]

5   Conclusion

In this chapter, we reviewed control approaches for reconfigurable robot manipulators. First, we introduced reconfigurable robot manipulators and their applications, putting emphasis on micro-scale ones. We classified the modular reconfigurable robot manipulators into three groups: reconfigurable serial, reconfigurable parallel, and self-reconfigurable robots. A review of existing (conventional, adaptive, and robust) control schemes for reconfigurable robot manipulators is provided, together with their benefits and drawbacks. Later, a focused review of specific control designs for micro-level robot manipulators is presented.

References

1. Abbott JJ, Nagy Z, Beyeler F, Nelson BJ (2007) Robotics in the small, part I: microrobotics. IEEE Robot Autom Mag 14(2):92–103
2. Ahmadzadeh H, Masehian E, Asadpour M (2015) Modular robotic systems: characteristics and applications. J Intell Robot Syst 81(3):317–357
3. Ambrose RO (1991) Design, construction and demonstration of modular, reconfigurable robots. Ph.D. thesis, University of Texas at Austin
4. Armstrong B, Khatib O, Burdick J (1986) The explicit dynamic model and inertial parameters of the Puma 560 arm. In: IEEE international conference robotics and automation, vol 1, pp 510–518
5. Asano H, Qiu ZJ, Zhou LB, Ojima H, Shimizu J, Ishikawa T, Edal H (2007) Path control scheme for vision guided micro manipulation system. Towards Synth Micro-Nano-Syst 2007(5):321–322
6. Astrom KJ (1983) Theory and applications of adaptive control-A survey. Automatica 19(5):471–486
7. Basha MA, Dechev N, Safavi-Naeini S, Chaudhuri SK (2007) Improved design of large 3-D micromirrors for microassembly into an optical MEMS cross-connect. Optomechatronic micro/nano devices and components III 2007, vol 6717, pp 71701–71701
8. Behkam B, Sitti M (2007) Bacterial flagella-based propulsion and on/off motion control of microscale objects. Appl Phys Lett 90, 023902
9. Bergeles C, Kratochvil BE, Nelson BJ (2012) Visually serving magnetic intraocular microdevices. IEEE Trans Robot 28(4):798–809

10. Beyeler F, Neild A, Oberti S, Bell DJ, Sun Y, Dual J, Nelson BJ (2007) Monolithically fabricated microgripper with integrated force sensor for manipulating microobjects and biological cells aligned in an ultrasonic field. J Microelectromech Syst 16(1):7–15

11. Bi ZM, Lang SYT (2005) General modular robot architecture and configuration design. In: IEEE international conference on mechatronics and automation, Niagara Falls, pp 268–273

12. Biglarbegian M, Melek W, Mendel JM (2011) Design of novel interval type-2 fuzzy controllers for modular and reconfigurable robots: theory and experiments. IEEE Trans Ind Electron 58(4):1371–1384

13. Bilen H (2008) Novel estimation and control techniques in micromanioulation using vision and force feedback. MASc thesis

14. Bouchebout S, Bolopion A, Abrahamians J-O, Regnier S (2012) An overview of multiple DoF magnetic actuated micro-robots. J Micro-Nano Mechatron 7(4):97–113

15. Brunete A, Hernando M, Gambao E (2005) Modular multi-configurable architecture for low diameter pipe inspection microrobots. In: IEEE international conference on robotics and automation, Barcelona, pp 490–495

16. Bryngelson R, Tosunoglu S (1994) On the design of a seven-axis modular robot. In: International conference on industrial electronics society (IECON), pp 1501–1506

17. Carbonari L, Callegari M, Palmieri G, Palpacelli MC (2014) A new class of reconfigurable parallel kinematic machines. Mech Mach Theory 79:173–183

18. Cavalcanti A, Shirinzadeh B, Freitas RA, Hogg T (2007) Nanorobot architecture for medical target identification. Nanotechnology 19(1), 015103

19. Chen I-M (1994) Theory and applications of modular reconfigurable robotic system. Ph.D. thesis, California Institute of Technology

20. Chen I-M (2001) Rapid response manufacturing through a rapidly reconfigurable robotic workcell. Robot Comput Integr Manuf 17:199–213

21. Chen I-M, Yang G (1996) Configuration independent kinematics for modular robots. In: IEEE international conference robotics and automation, Minneapolis, MN, pp 1440–1445

22. Chen W, Yang G, Ho EHL, Chen I-M (2003) Interactive-motion control of modular reconfigurable manipulators. In: International conference on intelligent robots and systems, Las Vegas, Nevada

23. Chen I-M, Yang G, Yeo S-H (2006) Automatic modeling for modular reconfigurable robotic systems, theory and practice, industrial-robotics-theory-modeling-control, Germany, p 964. ISBN 3-86611-285-8

24. Choi J-K, Omata T, Mori O (2004) Self-reconfigurable planar parallel robot. In: IEEE/RSJ international conference on intelligent robots and systems, Sendai, pp 2654–2660

25. Christensen DJ, Schultz UP, Stoy K (2013) A distributed and morphology-independent strategy for adaptive locomotion in self-reconfigurable modular robots. Robot Auton Syst 61(9):1021–1035

26. Chung WK, Han J, Youm Y, Kim S (1997) Task based design of modular robot manipulator using efficient genetic algorithm. In: IEEE international conference on robotics and automation, pp 507–512

27. Cohen R, Lipton MG, Dai MQ, Benhabib B (1992) Conceptual design of a modular robot. ASME J Mech Des 114:117–125

28. Coppola G, Zhang D, Liu K (2013) A 6-DOF reconfigurable hybrid parallel manipulator. Robot Comput Integr Manuf 30(2014):99–106

29. Craig JJ, Hsu P, Sastry SS (1986) Adaptive control of mechanical manipulators. In: IEEE international conference on robotics and automation, vol 3, pp 190–195

30. Dash AK, Chen I-M, Yeo SH, Yang G (2005) Task-oriented configuration design for reconfigurable parallel manipulator systems. Int J Comput Integr Manuf 18(7):615–634

31. Diller E, Pawashe C, Floyd S, Sitti M (2011) Assembly and disassembly of magnetic mobile micro-robots towards deterministic 2-D reconfigurable micro-systems. Int J Robot Res, 0278364911416140
32. Diller E, Floyd S, Pawashe C, Sitti M (2012) Control of multiple heterogeneous magnetic microrobots in two dimensions on nonspecialized surfaces. IEEE Trans Robot 28(1):172–182
33. Diller E, Zhang N, Sitti M (2013) Modular micro-robotic assembly through magnetic actuation and thermal bonding. J Micro-Bio Robot 8:121–131. doi:10.1007/s12213-013-0071-7
34. Donald B, Levey C, McGray C, Paprotny I, Rus D (2006) An untethered, electrostatic, globally controllable MEMS micro-robot. J Microelectromech Syst 15:1–15
35. Dong B, Li Y (2013) Decentralized reinforcement learning robust optimal tracking control for time varying constrained reconcilable modular robot based on ACI and function. Prob Eng 115(3):472–482
36. Duff DG, Yim M, Roufas K (2001) Evolution of polybot: a modular reconfigurable robot. In: Proceedings of the harmonic drive international symposium, Nagano
37. Ergeneman O, Dogangil G, Kummer MP, Abbott JJ, Nazeeruddin MK, Nelson BJ (2008) IEEE Sens J 8:20–22
38. Fisher R, Podhorodeski RP (2004) Design of a reconfigurable planar parallel manipulator. J Robot Syst 21(12):665–675
39. Freitas RA (2010) The future of nanomedicine. Futurist 44(1):21–22
40. Fukuda T, Kawauchi Y (1990) Cellular robotic system (CEBOT) as one of the realization of self-organizing intelligent universal manipulator. In: IEEE conference on robotics and automation, pp 662–667
41. Fukuda T, Nakagawa S (1988) Approach to dynamically reconfigurable robotic system. In: IEEE international conference on robotics and automation (ICRA), pp 1581–1586
42. Fukuda T, Nakagawa S, Kawauchi Y, Buss M (1988) Self organizing robots based on cell structures-CEBOT. In: IEEE international workshop on intelligent robots, pp 145–150
43. Gao Z, Zhang D, Ge Y (2010) Design optimization of a spatial six degree-of-freedom parallel manipulator based on artificial intelligence approaches. Robot Comput Integr Manuf 26(2):180–189
44. Gauthier M, Piat E (2006) Control of a particular micro-macro positioning system applied to cell micromanipulation. IEEE Trans Autom Sci Eng 3(3):264–271
45. Gavel DT, Siljak DD (1989) Decentralized adaptive control: structural conditions for stability. IEEE Trans Autom Control 34(4):413–426
46. Ghosh A, Fischer P (2009) Controlled propulsion of artificial magnetic nanostructured propellers. Nano Lett 9:2243–2245
47. Gilpin K, Kotay K, Rus D (2007) Miche Modular shape formation by self-disassembly. In: Proceedings of the 2007 IEEE international conference on robotics and automation (ICRA), Rome, pp 2241–2247
48. Gorman JJ, Dagalakis NG, Boone BG (2004) Multi-loop control of a nanopositioning mechanism for ultra-precision beam steering. Free-space laser communication and active laser illumination. SPIE, Bellingham, vol 5160, pp 170–181
49. Green P et al (1995) Demonstration of three-dimensional microstructure self-assembly. J Microelectoromech Syst 4-4:170–176
50. Hensinger DM, Johnston GA, Hinman-Sweeney EM, Feddema J, Eskridge S (2002) Self-reconfigurable robots. Sandia National Laboratories
51. Hosokawa K et al (1994) Dynamics of self-assembling systems: analogy with chemical kinetics. Artif Life 1(4):413–427
52. Hsu SH, Fu LC (2006) A fully adaptive decentralized control of robot manipulators. Automatica 42(10):1761–1767
53. Hsu P, Bodson M, Sastry S, Paden B (1987) Adaptive identification and control for manipulators without using joint accelerations. In: IEEE international conference on robotics and automation, vol 4, pp 1210–1215

54. Hui R, Kircanski N, Goldenberg A, Zhou C, Kuzan P, Wiercienski J, Gershon D, Sinha P (1993) Design of the iris facility-a modular, reconfigurable and expandable robot test bed. In: IEEE international conference on robotics and automation (ICRA), pp 155–160
55. Hun Q, Xu L, Zhang A (2012) Adaptive backstepping trajectory tracking control of robot manipulator. J Frankl Inst 349:1087–1105
56. Hunter IW, Jones LA, Sagar MA, Lafontaine SR, Hunter PJ (1995) Ophthalmic microsurgical robot and associated virtual environment. Comput Biol Med 25(2):173–177
57. Ioannou PA (1986) Decentralized adaptive control of interconnected systems. IEEE Trans Autom Control 31(4):291–298
58. Ioannou P, Fidan B (2006) Adaptive control tutorial. SIAM, Philadelphia
59. Ishiyama K, Arai K, Sendoh M, Yamazaki A (2000) Spiral-type micro-machine for medical applications. In: Proceedings of the international symposium on micromechatronics and human science (MHS), Nagoya, 22–25 October 2000, pp 65–69
60. Jantapremjit P, Austin D (2001) Design of a modular self-reconfigurable robot. In: Australian conference on robotics and automation, Sydney, pp 38–43
61. Jorgensen MW, Ostergaard EH, Lund HH (2004) Modular ATRON: modules for a self-reconfigurable robot. In: IEEE/RSJ international conference on intelligent robots and systems, vol 2, pp 2068–2073
62. Kamimura A, Kurokawa H, Yoshida E (2005) Automatic locomotion design and experiments for a modular robotic system. IEEE/ASME Trans Mechatron 10(3):314–325
63. Kasaya T, Miyazaki H, Saito S, Sato T (1999) Micro object handling under SEM by vision-based automatic control. In: IEEE international conference on robotics and automation, Detroit, pp 2189–2196
64. Kelly R, Santibanez V, Loria A (2005) Control of robot manipulators in joint space. Springer, London
65. Khalil I, Pichel M, Abelmann L, Misra S (2013) Closed-loop control of magnetotactic bacteria. Int J Rob Res 32:637–649
66. Kong X (2014) Reconfiguration analysis of a 3-DOF parallel mechanism using Euler parameter quaternions and algebraic geometry method. Mech Mach Theory 74:188–201
67. Kummer M, Abbott J, Kratochvil B, Borer R, Sengul A, Nelson B (2010) OctoMag: an electromagnetic system for 5-DOF wireless micromanipulation. IEEE Trans Robot 26:1006–1017
68. Kurokawa H et al (1998) A 3-D self-reconfigurable structure and experiments. In: IEEE/RSJ international conference on intelligent robots and systems, pp 860–865
69. Li W, Slotine J-JE (1988) A rapidly deployable manipulator system. In: IEEE international conference on robotics and automation, vol 2, pp 704–709
70. Li Z, Melek W, Clark CM (2007) Development and characterization of a modular and reconfigurable robot. In: The 2nd international conference on changeable, agile, reconfigurable and virtual production (CARV 2007), Toronto, 22–24 July 2007
71. Li Y, Zhu M, Li Y (2008) Velocity-observer-based compensator for motion control of a reconfigurable manipulator. Control Theory Appl 25(5):891–897
72. Li Z, Melek WW, Clark C (2009) Decentralized robust control of robot manipulators with harmonic drive transmission and application to modular and reconfigurable serial arms. Robotica 27(2):291–302
73. Li Y, Zhu M, Li Y (2009) Observer-based decentralized adaptive fuzzy control for reconfigurable manipulator. Control Decis 24(3):429–434
74. Liu G, Abdul S, Goldenberg AA (2006) Distributed modular and reconfigurable robot control with torque sensing. In: IEEE international conference on mechatronics and automation, pp 384–389
75. Liu G, Abdul S, Goldenberg AA (2008) Decentralized robust control of modular and reconfigurable robots with harmonic drive. Robotica 26(1):75–84
76. Liu G, Liu Y, Goldenberg AA (2011) Design, analysis, and control of a spring-assisted modular and reconfigurable robot. IEEE/ASME Trans Mechatron 16(4):695–706
77. Madhevan B, Sreekumar M (2012) Structures and characteristics in reconfigurable modular robots. Advances in reconfigurable mechanisms and robots I. Springer, London, 525–534

78. Mahoney A, Sarrazin J, Bamberg E, Abbott J (2011) Velocity control with gravity compensation for magnetic helical microswimmers. Adv Robot 25:1007–1028
79. Marino H, Bergeles C, Nelson BJ (2014) Robust electromagnetic control of microrobots under force and localization uncertainties. IEEE Trans Autom Sci Eng 11(1)
80. Martel S, Felfoul O, Mathieu JB, Chanu A, Tamaz S, Mohammadi M, Mankiewicz M, Tabatabaei N (2009) Int J Robot Res 28:11–69
81. Mastrangeli M et al (2009) Self-assembly from milli-to nanoscales: methods and applications. J Micromech Microeng 19(8):083001
82. Matsumaru T (1995) Design and control of the modular robot system: TOMMS. In: IEEE international conference on robotics and automation, vol 2, pp 2125–2131
83. Melek WW, Goldenberg AA (2003) Neurofuzzy control of modular and reconfigurable robots. IEEE/ASME Trans Mechatron 8(3):381–389
84. Merlet J-P (2006) Parallel robots, 2nd edn. Springer/Hermes, Paris
85. Middleton RH, Goodwin GC, Mendel JM (1988) Adaptive computed torque control for rigid link manipulations. Syst Control Lett 10(1):9–16
86. Mobes S, Laurent GJ, Clevy C, Le Fort-Piat N, Piranda B, Bourgeois J (2012) Toward a 2D modular and self-reconfigurable robot for conveying microparts, second workshop on design, control and software implementation for distributed MEMS
87. Murata S, Kurokawa H (2007) Self-reconfigurable robots. IEEE Robot Autom Mag 14(1):71–78
88. Murata S, Kurokawa H, Kokaji S (1994) Self-assembling machine. In: Proceedings of IEEE international conference on robotics and automation, vol 1, pp 441–448
89. Murata S et al (1998) A 3-D self-reconfigurable structure. In: Proceedings of IEEE international conference on robotics and automation, pp 432–439
90. Murata S, Yoshida E, Kamimura A, Kurokawa H, Tomita K, Kokaji S (2002) M-TRAN: self-reconfigurable modular robotic system. IEEE/ASME Trans Mechatron 7(4)
91. Nagy PV (1988) The PUMA 560 industrial robot: inside-out. Creative Manufacturing Engineering Program, MS88-280. Society of Manufacturing Engineers, Dearborn, pp 4/67-4/79
92. Padhy SK (1992) On the dynamics of SCARA robot. Robot Autom Syst 10:71–78
93. Palpacelli M-C, Carbonari L, Palmieri G, Callegari M (2015) Analysis and design of a reconfigurable 3-DoF parallel manipulator for multimodal tasks. IEEE/ASME Trans Mechatron 20(4):1975–1985
94. Pamecha A, Chirikjian GS (1996) A useful metric for modular robot motion planning. In: IEEE international conference on robotics and automation, pp 442–447
95. Pamecha A, Ebert-Uphoff I, Chirikjian GS (1997) Useful metrics for modular robot motion planning. IEEE/ASME Trans Robot Autom 10(3):314–325
96. Paprotny I, Bergbreiter S (2013/2014) Small-scale robotics. LNAI, vol 8336. Springer, Berlin/Heidelberg, pp 1–15
97. Paredis CJJ, Brown HB, Khosla PK (1996) A rapidly deployable manipulator system. In: IEEE international conference on robotics and automation, vol 2, pp 1434–1439
98. Pawashe C, Floyd S, Sitti M (2009) Modeling and experimental characterization of an untethered magnetic micro-robot. Int J Robot Res 28(8):1077–1094
99. Pawashe C, Floyd S, Sitti M (2011) Assembly and disassembly of magnetic mobile micro-robots towards 2-D reconfigurable micro-systems. In: IEEE international conference on robotics and automation (ICRA), Shanghai, pp 261–266
100. Robot Institute of America (1979)
101. Rus D, Vona M (2000) A physical implementation of the self-reconfiguring crystalline robot. In: Proceedings of the 2000 IEEE international conference on robotics and automation (ICRA), San Francisco, pp 1726–1733
102. Rus D, Vona M (2001) Crystalline robots: self-reconfiguration with compressible unit modules. J. Auton Robot 10(1):107–124
103. Schmitz PKD, Kanade T (1988) The CMU reconfigurable modular manipulator system. Technical report CMU-RI-TR-88-07, Robotics Institute, Pittsburgh, PA

104. Schweikardt E Current trends and miniaturization challenges for modular self-reconfigurable robotics. Carnegie Mellon University, Pittsburgh
105. Shahini M, Melek WW, Yeow JTW (2009) Micro-force compensation in automated micro-object positioning using adaptive neural networks. Smart Mater Struct 18:095023 (14pp)
106. Shahini M, Melek WW, Yeow JTW (2010) Characterization of surface micro forces under varying operational conditions in micro-sized object pushing: an empirical approach. J Micromech Microeng 20(5):055013. Published 14 April 2010
107. Shahini M, Melek WW, Yeow JTW (2013) Automated sequential pushing of micro objects by using adaptive controller. Int J Robot Autom 2(4):163
108. Shen YT, Xi N, Wejinya UC, Li WJ, Xiao JH (2004) Infinite dimension system approach for hybrid force/position control in micromanipulation. In: Proceedings of IEEE international conference on robotics and automation, vols 1–5, pp 2912–2917
109. Shen WM, Chiu H, Rubenstein M, Salemi B (2008) In: Proceedings of the space technology international forum, Albuquerque, New Mexico
110. Shi L, Singh SK (1992) Decentralized adaptive controller design for large-scale systems with higher order interconnections. IEEE Trans Autom Control 37(8):1106–1118
111. Siljak DD (1991) Decentralized control of complex systems. Mathematics in science and engineering, vol 184. Dover, New York
112. Sitti M (2000) Controlled pushing of nanoparticles: modeling and experiments. IEEE/ASME Trans. Mechatron. 5(2):199–221
113. Sitti M (2004) Atomic force microscope probe based controlled pushing for nanotribological characterization. IEEE-ASME Trans Mechatron 9(2):343–349
114. Sitti M (2007) Microscale and nanoscale robotics systems [grand challenges of robotics]. IEEE Robot Autom Mag 14(1):53–60
115. Sitti M (2009) Voyage of the microrobots. Nature 458:1121–1122
116. Slotine J-JE (1985) The robust control of robot manipulators. Int J Robot Res 4(2):49–64
117. Slotine J-JE, Li W (1986) Adaptive control of mechanical manipulators. In: Winter annual meeting (ASME)
118. Slotine J-JE, Li W (1988) Adaptive manipulator control: a case study. IEEE Trans Automat Control 33(11):995–1003
119. Spong MW, Thorp JS, MJ (1987) Robust microprocessor control of robot manipulators. Automatica 23(3):373–379
120. Su CY, Leung TP (1993) A sliding mode controller with bound estimation for robot manipulators. IEEE Trans Robot Autom 9(2):208–214
121. Suh JW, Homans SB, Yim M (2002) Telecubes: mechanical design of a module for self-reconfigurable robotics. In: Proceedings of the IEEE international conference on robotics and automation (ICRA), pp 4095–4101
122. Sul O, Falvo M, Taylor R, Washburn S, Superfine R (2006) Thermally actuated untethered impact-driven locomotive microdevices. Appl Phys Lett 89:203512
123. Sun Y, Nelson BJ (2002) Biological cell injection using an autonomous microrobotic system. Int J Robot Res 21(10–11):861–868
124. Sun D, Hu S, Shao X, Liu C (2009) Global stability of a saturated nonlinear PID controller for robot manipulators. IEEE Trans Ind Electron 17(4):892–899
125. Tang Y, Tomizuka M, Guerrero G, Montemayor G (2006) Decentralized robust control of mechanical systems. IEEE Trans Autom Control 45(4):771–776
126. Tarokh M (1999) Decoupled nonlinear three-term controllers for robot trajectory tracking. IEEE Trans Robot Autom 15(2):369–380
127. Thompson JA, Fearing RS (2001) Automating microassembly with ortho-tweezers and force sensing. In: Proceedings of the 2001 IEEE/RJS/IROS international conference on intelligent robots and systems, vols 1–4, pp 1327–1334
128. Tomita K, Murata S, Kurokawa H, Yoshida E, Kokaji S (1999) Self assembly and self-repair method for a distributed mechanical system. IEEE Trans Robot Autom pp 1035–1045
129. Tomita K et al (1999) Self-assembly and self-repair method for distributed mechanical system. IEEE Trans Robot Autom 15(6):1035–1045

130. Unsal C et al (1999) I(CDS)-cubes: a modular self-reconfigurable bipartite robotic system. In: Proceedings of SPIE, sensor fusion and decentralized control in robotic systems II, pp 258–269

131. Vollmers K, Frutiger D, Kratochvil B, Nelson B (2008) Wireless resonant magnetic microactuator for untethered mobile microrobots. Appl Phys Lett 92:144103

132. White PJ, Kopanski K, Lipson H (2004) Stochastic self-reconfigurable cellular robotics. In: Proceedings of the 2004 IEEE international conference on robotics and automation (ICRA), New Orleans, pp 2888–2893

133. Whitesides GM, Grzybowski B (2002) Self-assembly at all scales. Science 295(5564):2418–2421

134. Xu T, Hwang G, Andreff N, Regnier S (2015) Planar path following of 3-D steering scaled-up helical microswimmers. IEEE Trans Robot 31:117–127

135. Yang G, Chen I-M, Lim WK, Yeo SH (1999) Design and kinematic analysis of modular reconfigurable parallel robots. In: IEEE international conference on robotics and automation, Detroit, MI, pp 2501–2506

136. Yang G, Chen IM, Lim WK, Yeo SH (2001) Kinematic design of modular reconfigurable in-parallel robots. Autonomous robots, vol 10, pp 83–89. Kluwer Academic Publishers, Manufactured in The Netherlands

137. Yang G, Chen I-M, Lim WK, Yeo SH (2001) Self-calibration of three-legged modular reconfigurable parallel robots based on leg-end distance errors. Robotica, vol 19. Cambridge University Press, Cambridge, pp 187–198. (Printed in the United Kingdom)

138. Yang G, Chen I-M, Chen W, Yeo SH (2003) Design and analysis of a 3-RPRS modular parallel manipulator for rapid deployment. In: International conference on advanced intelligent mechatronics (AIM 2003)

139. Yang G, Chen I-M, Chen W, Yeo SH (2003) Design and analysis of a 3-RPRS modular parallel manipulator for rapid deployment. In: IEEE/ASME international conference on advanced intelligent mechatronics, pp 1250–1255

140. Ye W, Fang Y, Zhang K, Guo S (2014) A new family of reconfigurable parallel mechanisms with diamond kinematotropic chain. Mech Mach Theory 74:1–9 (2014)

141. Yesin KB, Vollmers K, Nelson BJ (2006) Modeling and control of untethered biomicrorobots in fluid environment using electromagnetic fields. Int J Robot Res 25(5/6):527–536

142. Yim M, Duff DG, Roufas KD (2000) PolyBot: a modular reconfigurable robot. In: Proceedings of the 2000 IEEE international conference on robotics and automation (ICRA), San Francisco, pp 514–520

143. Yim M, Shirmohammadi B, Sastra J, Park M, Dugan M (2007) Towards robotic self-reassembly after explosion. In: IEEE/RSJ international conference on intelligent robots and systems, San Diego, CA, Oct 29–Nov 2 2007

144. Yim M, Shen WM, Salemi B, Rus D, Moll M, Lipson H, Klavins E, Chirikjian GS (2007) Modular self-reconfigurable robot systems. IEEE Robot Autom Mag 14(1):43–52

145. Yoshida E, Murata S, Kokaji S, Tomita K, Kurokawa H (2000) Micro self-reconfigurable robotic system using shape memory alloy. Distributed autonomous robotic systems, vol 4. Springer, Berlin, pp 145–154

146. Yoshida E, Kokaji S, Murata S, Tomita K, Kurokawa H (2001) Miniaturized self-reconfigurable system using shape memory alloy. J Robot Mechatron 13:212

147. Yoshida E, Murata S, Kamimura A, Tomita K, Kurokawa H, Kokaji S (2001) Motion planning for a self-reconfigurable modular robot. In: Experimental robotics VII. Springer, Berlin/Heidelberg, pp 385–394

148. Zhang D (2010) Parallel robotic machine tools. Springer, New York

149. Zhang D, Bi Z (2006) Development of reconfigurable parallel kinematic machines using modular design approach. In: CDEN/RCCI international design conference, University of Toronto, ON

150. Zhang D, Gao Z (2012) Forward kinematics, performance analysis, and multi-objective optimization of a bio-inspired parallel manipulator. Robot Comput-Integr Manuf 28(4):484–492

151. Zhang L, Abbott J, Dong L, Kratochvil B, Bell D, Nelson B (2009) Artificial bacterial flagella: fabrication and magnetic control. Appl Phys Lett 94:064107
152. Zhao B, Li Y (2014) Decentralized differential tracker based control of reconfigurable manipulators without velocity sensor. In: 33rd Chinese control conference (CCC), pp 8370–8374
153. Zhao J, Ren Z, Zhang Y (2006) Configuration matching in self-reconfigurable process of modular self-reconfigurable robots. In: IEEE international conference on mechatronics and automation, Luoyang, pp 284–288
154. Zhu WH, Lamarche T (2007) Modular robot manipulators based on virtual decomposition control. In: IEEE international conference on robotics and automation (ICRA). IEEE, New York, pp 2235–2240
155. Zhu M, Li Y (2010) Decentralized adaptive fuzzy sliding mode control for reconfigurable modular manipulators. Int J Robust Nonlinear Control 20(4):472–488
156. Zeng Q, Ehmann KF (2014) Design of parallel hybrid-loop manipulators with kinematotropic property and deployability. Mech Mach Theory 71:1–26 (2014)

Chapter 10
MEMS Sensors and Actuators

Vishwas N. Bedekar and Khalid Hasan Tantawi

Abstract Over the past 30 years, significant progress has been made in the fields of Micro-Electro-Mechanical Systems (MEMS). Originally, the term MEMS was used in the United States to signify electromechanical sensors and actuators that are based on semiconductors. In Europe, the same technology was referred to as "Microsystem Technology," while in Japan the term micromechatronics was used. Today, the term MEMS refers to all devices, sensors, and actuators that have dimensions in the micrometer range, whether based on semiconductors or other materials such as glass and metals. In this chapter a brief review of the history is provided followed by future trends in the Micro-electro-mechanical Systems (MEMS) industry. We further present materials that are commonly used in the fabrication of MEMS sensors and actuators and we discuss the advantages and disadvantages of each material. Finally, the different types, applications, and recent technologies in MEMS sensors and actuators constitute the core of this chapter.

Keywords Microelectromechanical systems • Sensors • Actuators • Materials

1 Introduction to MEMS Sensors and Actuators

Over the past 30 years, significant progress has been made in the fields of Micro-Electro-Mechanical Systems (MEMS). Originally, the term MEMS was used in the United States to signify electromechanical sensors and actuators that are based on semiconductors. In Europe, the same technology was referred to as "Microsystem Technology," while in Japan the term micro-mechatronics was used. Today, the term MEMS refers to all devices, sensors, and actuators that have dimensions in the micrometer range, whether based on semiconductors or other materials such as

V.N. Bedekar (✉)
Department of Engineering Technology, Middle Tennessee State University,
Murfreesboro, TN 37132, USA
e-mail: Vishwas.Bedekar@mtsu.edu

K.H. Tantawi
Department of Career Readiness-Mechatronics, Motlow State Community College,
Smyrna, TN 37167, USA

glass and metals [1]. MEMS sensors and actuators are usually classified by the field in which they are used. For instance, MEMS for radio frequency applications are referred to as RF-MEMS, in the same way optical MEMS (also known as Micro-Opto-Electro-Mechanical Systems MOEMS) and BioMEMS refer to devices in their specific applications.

Sensors that are used in the smart phone industry such as fingerprint, motion, gyroscope, and pressure sensors have seen a significant improvement in the recent years propelled by the high demand of smart phones and tablets. It is estimated that the revenue of fingerprint sensors will quadruple in the year 2020 compared to that in 2014 to peak at $1.75 billion [2].

On the other hand, the powertrain industry has been driving the advances in MEMS sensors that are used in internal combustion engines, such as temperature, pressure monitoring, position, acceleration, oxygen, and knock sensors. Analysts at the market research company IHS Technology expect that shipments of sensors that are used in internal combustion engines will increase by 2.6 % to 1.36 billion units in the year 2021 compared to that in 2016 [3]. The top 20 manufacturers of MEMS sensors and actuators such as Bosch, STMicroelectronics, Knowels, and Freescale claimed a combined revenue in excess of $7 billion in 2013 [4].

MEMS sensors and actuators found significant applications in the medical industry as well; applications include optical image sensing and miniaturized robots for minimum invasive surgical procedures as well as drug delivery. A picture of a system that utilizes microrobotics equipped with numerous MEMS sensors and actuators is shown in Fig. 10.1.
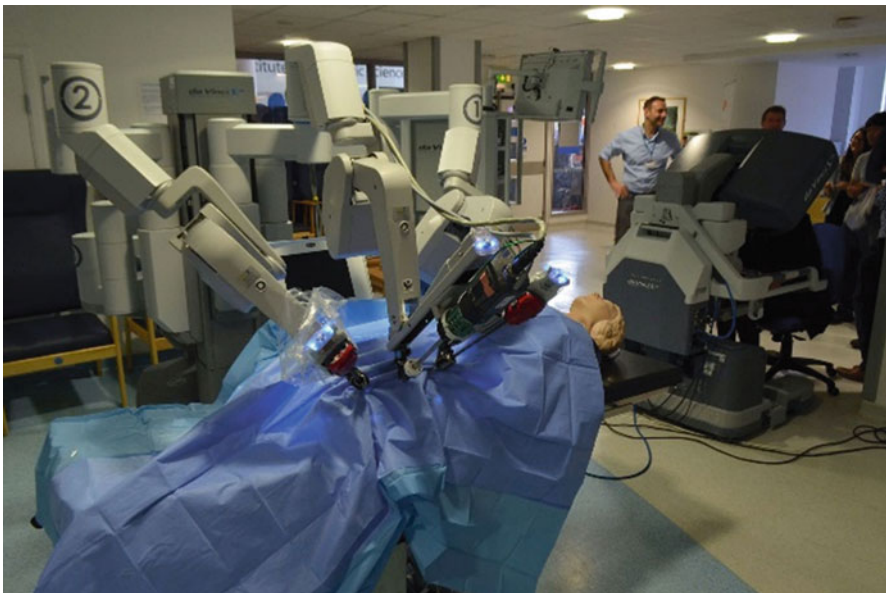

Fig. 10.1 Many MEMS sensors and actuators are found in surgical microrobotics such as those in the Da Vinci surgical system that is shown in the picture, allowing for minimally invasive surgical procedures [5]

Since the turn of the twenty-first century, biology has been one of the main fields where MEMS applications have excelled, that a specific term is used, BioMEMS. For example, semiconductor-based structures have been used in drug delivery [7]. Biosensors with ultra-high sensitivity, such as DNA and protein strand sensors [8] can be fabricated from porous materials such as porous silicon and porous carbon [9, 10].

2   Materials for MEMS

Different materials have been used for MEMS processing, such as silicon, quartz, photodefinable glasses, metals, polymers, and ceramics. Each material has its own distinct features and limitations. Silicon is probably the most commonly used material in MEMS industry.

2.1   Silicon

Silicon is a semiconductor material that belongs to group 14 in the periodic table of elements. It has four features that make it stand out over other materials for many MEMS applications:

1. Silicon's semiconductive property allows for developing numerous electronic devices on the same substrate. The electrical conductivity of silicon can be easily controlled by doping silicon. Furthermore, silicon can be easily made electrically isolative, by oxidizing it. The semiconductive property of silicon is what makes it the preferred material for the microelectronics industry.
2. Silicon's biocompatibility: this property allows silicon to be used in equipment and devices that are used in invasive medical procedures, without being rejected by the biological immune system. In addition to that, silicon is biodegradable to a nontoxic material (orthosilicic acid Si(OH)4) [11].
3. Silicon's optical properties: The photoluminescence property and photovoltaic effect of silicon allow it to be the preferred material for applications that involve photosensing, or monitoring based on photoemission, such as in drug delivery, in which *in-vivo* monitoring of the drug cargo being delivered in a biological body is needed [12].
4. The wide availability and low cost of silicon: This is one of its main features; silicon is one of the most common materials in the earth.

2.2 Glass

Glasses of different types and compositions are also widely used materials in MEMS. The main features of glass over silicon and other materials include:

1. Optical transparency over a wide range of wavelengths.
2. Electrical Insulation
3. Low thermal conductivity
4. Chemical inertness

These features make glass the preferred choice in many applications, such as in microfluidics, and optical MEMS, where optical transparency over a wide range in the infrared and/or the ultraviolet/visible region is needed, as well as the necessity of handling chemicals, that may react with other materials. In addition to that, the low thermal expansion of glass makes it a suitable material for applications such as in piezoelectric devices.

Quartz is the most commonly used type of glass, due to its wide transparency, that ranges from about 200 nm to 4 μm. For Optical MEMS applications that require transparency in the mid-wave and long-wave infrared regions (wavelengths up to about 12 μm) such as in many IR detectors and optical microcavities, chalcogenide glasses are commonly used.

Another type of glass that has found potentials in MEMS is the glass referred to as photodefinable glass. This glass belongs to the Lithium-Aluminum-Silicate family [13] with impurities of metal oxides, in particular, cerium and silver that greatly contribute to the photostructurability of this family of glasses. A typical procedure to fabricate structures from photodefinable glasses is to expose the glass to mid-to-near ultraviolet (UV) light (280–310 nm) through a quartz mask. This results in a latent image of the mask in the glass matrix. The glass then undergoes a heat treatment stage in which silver nanoclusters are formed in the exposed regions that act as nuclei for crystallizing lithium metasilicate on them. In the next step, the exposed regions are etched in hydrofluoric (HF) acid. Finally, the etched substrate is thermally annealed to provide an optically transparent and scatter-free surface [14–17]. Influence of cerium and the effect of UV dose on Foturan™ photodefinable glass transmission in the UV/Visible and near-infrared (NIR) regions were shown by Bhardwaj [18] and Livingston et al. [19, 20]. The low cost and ability to mass produce microfluidic structures in photodefinable glass, in addition to the ability to fabricate structures with a high surface finish [4], make this technology a cheaper alternative for quartz-based technologies [5, 9, 14, 17, 21] in many applications. Figure 10.2 shows an optical micrograph of different structures made in photodefinable glass.

Quality of the glass surface finish is one of the important factors to consider for Optical MEMS that operate in the UV region. For this purpose, glass micromachining technologies such as glass etching [13–15] and laser drilling [16], offer high surface finish, but they are limited to heights of less than 160 μm.

Fig. 10.2 Microstructures fabricated in photodefinable glass
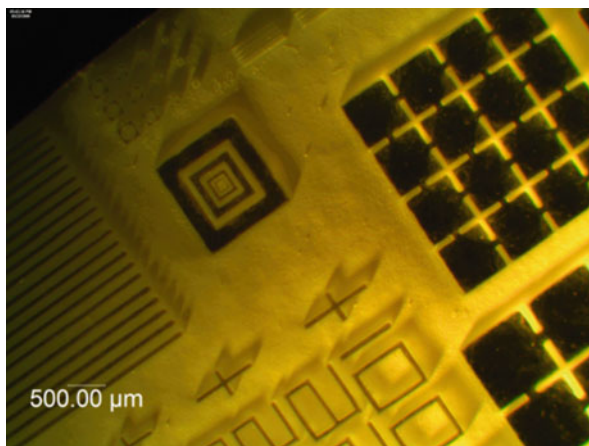
2.3 Polymers

Polymers such as Polydimethylsiloxane (PDMS) are very commonly used in fabricating MEMS devices. The main features of PDMS include its transparency to a wide range in the visible and infrared regions, low fabrication cost, as well as its elasticity. PDMS finds applications in the microfluidics industry, such as microchannels patterned on polymer or glass substrates [6, 10]. A PDMS-based micropump that is used for microfluidics is shown in Fig. 10.3. Despite the attractive features of PDMS, microfluidic devices that are PDMS-based suffer from high hydrophobicity [11], porosity [12], and solubility in many organic solvents, which significantly limit their usability. Furthermore, PDMS has limited optical transparency [22] windows when compared to Quartz and commercial photosensitive glasses.

2.4 Metals and Metal Alloys

A wide range of metals and their alloys are employed in sensor technologies for different applications ranging from highly sophisticated communication applications in RF-MEMS to applications in which they only provide mechanical and structural support for handling severe environments. Some commonly used metals in the MEMS industry include copper, gold, platinum, titanium, and tungsten.
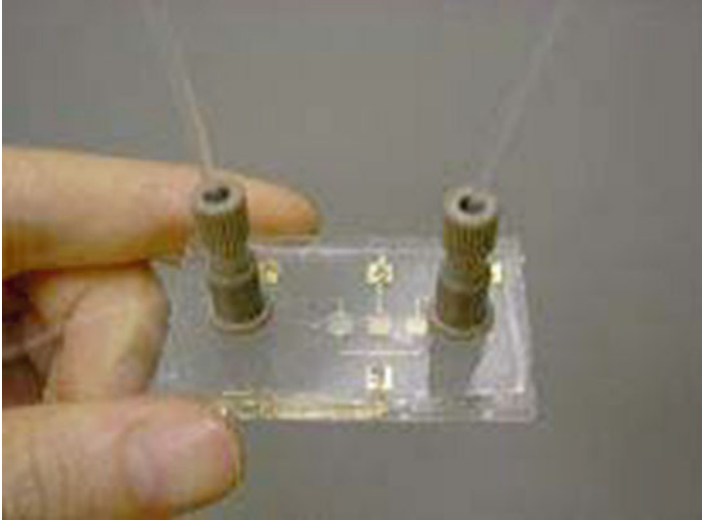
Fig. 10.3 A PDMS-based micropump [23]

2.5 Ceramics

Most common ceramics used for MEMS applications are silicon carbide, titanium
nitride, lead zirconate titanate (PZT), and aluminum nitride PZT is widely used
in MEMS sensor and actuator devices for the unique property of this material to
convert energy from one form to another. When used as a sensor or actuator, it
converts electric energy into mechanical strain using converse piezoelectric effect.
Polster et al. used pyramidal aluminum nitride 3-dimensional structures to detect
normal and shear forces.

3 MEMS Sensors

MEMS-based sensors for detecting chemical, physical, and biological quantities
can be found in numerous applications in our everyday lives. For instance, MEMS
accelerometers can be found in applications ranging from cameras and video game
consoles to aircraft attitude control systems. A scanning electron microscope (SEM)
image of an CMOS accelerometer is shown in Fig. 10.4.

Many principles of operation are employed in MEMS-based pressure sensors,
such as piezoresistive and the more accurate capacitive pressure sensors.

Semiconductor-based gas sensors have the advantages of being able to be
integrated with electric circuits. MEMS devices that are used for gas detection are
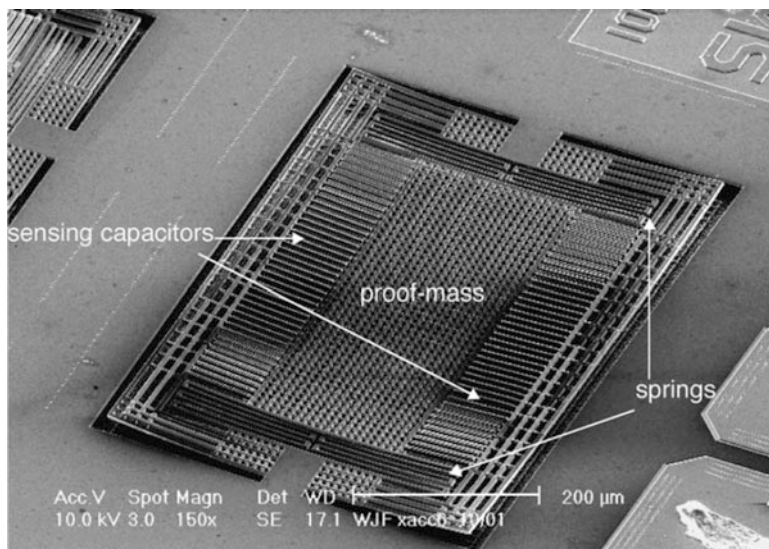usually made of ceramics, thin films of semiconductor-metal gas detectors.

Fig. 10.4 An accelerometer based on the CMOS MEMS technology [24]

Sensors that detect chemicals in the gaseous phase have many applications ranging from leakage detection of explosive and toxic gasses in an industrial setting to environmental monitor and control applications.

Criteria and issues that must be considered when fabricating MEMS sensors may include (but not limited to) [1, 25]:

- **Sensitivity**: This is the main criterion in a sensor, for sensors that detect chemical molecules, the sensitivity increases as the exposure area of the sensing material to the environment increases; therefore, sensitivity is enhanced as the surface area-to-volume ratio of the sensing material increases. The sensitivity corresponds to the lowest detectable change of the quantity or concentration of the chemical molecule that is being detected.
- **Selectivity**: the sensor needs to be selective to the physical quantity or the chemical molecule being sensed to avoid false triggers.
- **Biocompatibility**: sensors that involve detection of biomolecules such as glucose sensors must use biocompatible materials.
- **Real-time operation**: in most cases, the sensor must be able to operate within a time constraint; this is determined by the response time of the sensor. In addition to that, the sensor should be able to recover to its sensing state within a time constraint.
- **Temperature dependence**: most sensors are significantly affected by temperature; thus, the operating temperature range is an important criterion for sensors.
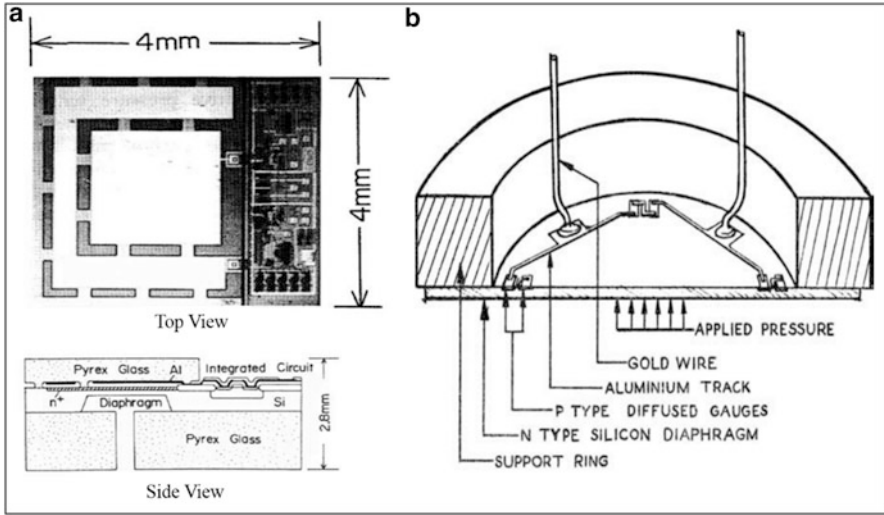
Fig. 10.5 (**a**) A modern MEMS pressure sensor shown in comparison to an older pressure sensor (**b**) that was manufactured in the 1960's [26]

- **Sensor stability in performance**: An important criterion for a sensor is that it should be stable and able to reproduce its results consistently in the ambient environment.

3.1 Sensors Based on Piezoelectric and Piezoresistive Effects

Piezoelectric materials experience a deformation in their crystals when a voltage is applied to them, or inversely, generate an electric charge when a mechanical stress is applied to them. This property allows these materials to be used in sensors and actuators of numerous applications such as in accelerometers, temperature, viscosity, and pressure sensors, and many other applications. On the other hand, piezoresistive materials exhibit a change in their electrical resistance when they undergo a mechanical strain. A device that is used for pressure sensing and based on piezoresistivity, with an integrated circuit and a non-planar diaphragm is shown in comparison to an older sensor in Fig. 10.5.

Another application in which piezoresistivity is employed is shown in the stress-contact sensor shown in Fig. 10.6 [27]. The sensor is composed of a diaphragm that bends with applied stresses, four resistors are connected to the 1000 μm—thick diaphragm to form the sensor.

A state-of-the-art sensor that resembles the human skin and uses piezoelectric and ferroelectric effects is shown in comparison to a fingertip skin structure in Fig. 10.7. This artificial skin sensor uses piezoelectric effects to sense contact with

Fig. 10.6 A micrograph of a stress-contact sensor with a US penny in its background to demonstrate its size (*left*). The sensor structure (*right*) [27]
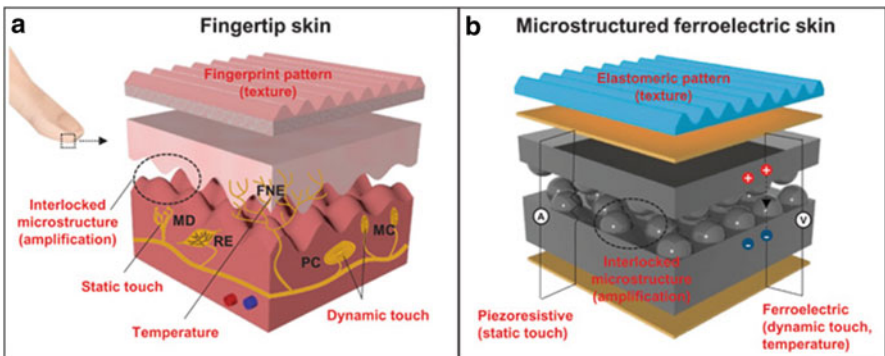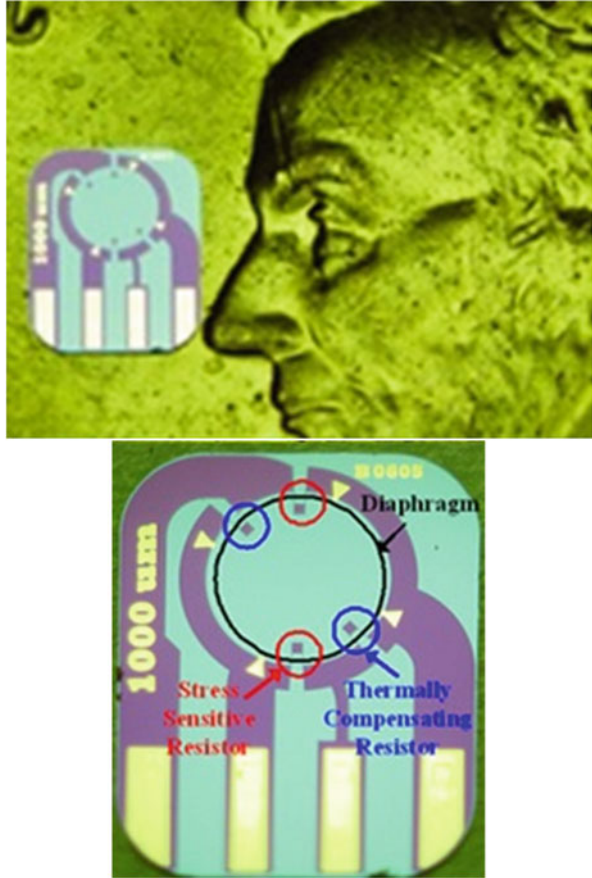

Fig. 10.7 An artificial skin that senses objects, texture, and temperature in comparison to the skin of a human fingertip [28, 29]

other surfaces and ferroelectric effects to detect surface roughness (texture) and temperature of the objects being sensed [28, 29].

3.2 Sensors Based on Thin Films

Thin-film gas sensors are very commonly used in different applications such as in alcohol sensors for breath analyzers and home CO and hydrocarbon sensors [30]. These gas sensors are made of a thin oxide film of a noble metal such as platinum, titanium, or zinc that is a few micrometers thick and deposited on a substrate that is usually of n-type silicon to form something similar to a Schottky diode. When the metal oxide is in a gaseous atmosphere, its electrical conductivity changes [31].

Gas sensitivity of thin-film sensors is achieved by the exposure of the surface of the solid thin film to the gas atmosphere; thus, these sensors have limited sensitivity due to the limitation in the surface areas. Significant improvement in the sensitivity can be achieved by using hollow or porous materials such as carbon nanotubes and porous silicon. These materials have the advantage of having extremely high surface-area-to-volume ratios, which allow them to be used for ultra-sensitive gas detection and in biosensing [9, 10].

3.3 Sensors Based on Carbon Nanotubes

Carbon nanotubes are fullerene structures composed of tubes of single or multi-carbon atom walls. The C–C bonds in these structures make them very strong and stable. In addition to that, their electrical conductivity can be controlled [25]. They have been demonstrated to have a high sensitivity to small gas molecules in room temperature, compared to temperatures of few hundred degrees Celsius that are required for thin-film sensors [32]. One process that is used to fabricate carbon nanotubes is referred to as the High Pressure Carbon Monoxide (HiPCO) process, which was developed by a team from Rice University and published in 2001 [33]. In this process, a reaction in the gaseous phase takes place between iron carbonyl $Fe(CO)_5$ and carbon monoxide (CO), forming iron clusters in the gas, which in turn serves as nucleation sites on which carbon nanotubes are grown.

Li et al. [32] developed a gas sensor in which a web of single-walled nanotubes stretches between two gold electrodes as shown in Fig. 10.8. The electrodes were fabricated by thermal evaporation of a 20 nm layer of titanium a 40 nm layer of gold on a thermally grown silicon dioxide. The carbon nanotubes were fabricated using the HiPCO process, and then purified to a level of 99.6 %. This gas sensor was demonstrated for NO_2, acetone, benzene, and nitrotoluene.

Fig. 10.8 A gas sensor made
of carbon nanotubes in
between gold electrodes [32]

Fig. 10.8 A gas sensor made of carbon nanotubes in between gold electrodes [32]
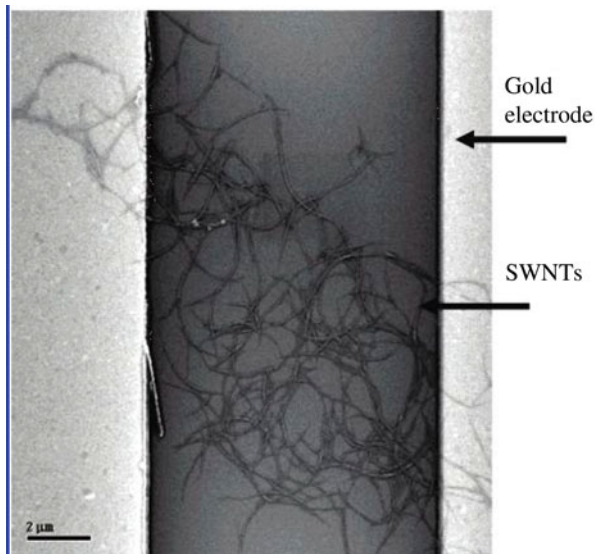
3.4   Sensors Based on Porous Materials

Other materials that show high potentials for detection of physical and biological quantities are porous structures such as porous silicon and porous alumina. Porous silicon was first discovered in 1956 by Uhlir through electrochemical etching of silicon. This discovery paved the way for devices based on quantum structures [1]. The high surface area-to-volume ratio of porous silicon makes it a suitable material for sensing humidity, gas detection, and detection and characterization of biological membranes. Two different porous silicon structures are shown in Fig. 10.9, a membrane of 3-μm thickness (shown on the left) [34] and a substrate of about 15 μm deep pores (shown on the right) [35].

MEMS structures based on porous silicon [36, 37] have a strong potential for applications that require detecting biological agents, mainly due to the biocompatibility of this material and its high surface area-to-volume ratio. Different porous silicon structures that were used to characterize and sense biological membranes and proteins are shown in Fig. 10.10. Proteins that reside in the cellular membranes are of particular importance for scientists due to the fact that they are the targets of most pharmaceutical drugs. However, the complex environment in which they reside, makes it difficult to characterize these proteins, many transmembrane proteins experience denaturing immediately once their native environment is changed. For this reason, technologies that are based on the biocompatible porous silicon were developed to sense and characterize biological membranes and transmembrane proteins such as the one shown by Worsfold et al. [38], which used porous silicon to support lipid bilayers that are used for biosensing human umbilical vein endothelial cells. This device was composed of a porous silicon structure with a coated lipid
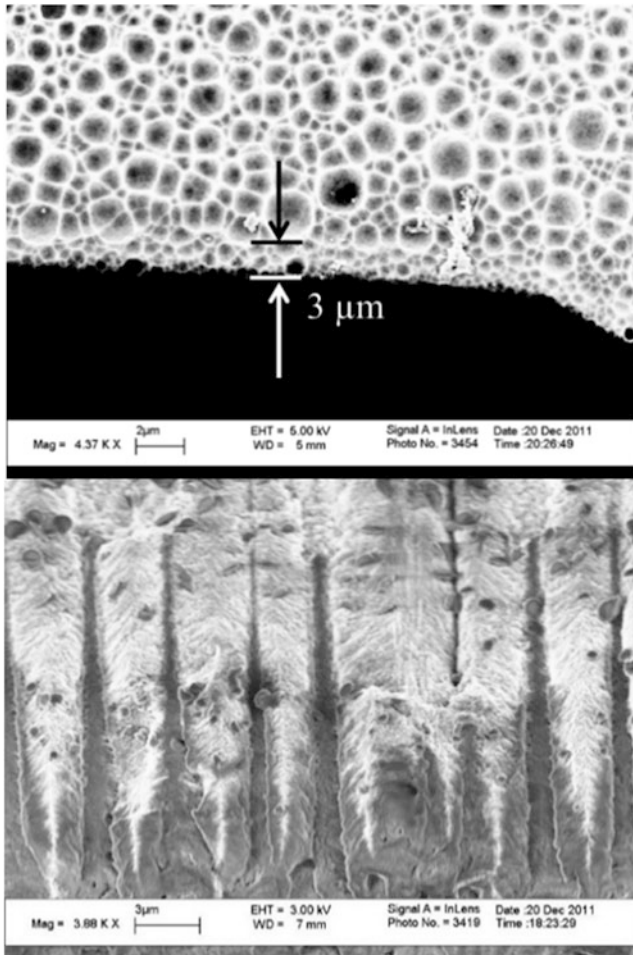
Fig. 10.9 *Left*—a 3 μm-thick porous silicon membrane used to support biological membranes [34]. *Right*—a porous silicon structure of pores that are about 1.5 μm wide and 15 μm deep [35]
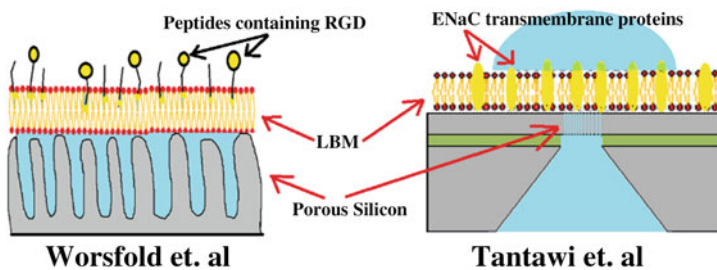

Fig. 10.10 Two different devices that are based on the biocompatible material porous silicon for detection and characterization of biological structures [35]

bilayer with arginine-glycine-aspartate that contained peptides of integrin $\alpha_v\beta_3$ on the porous silicon. The artificial cell membrane shown by Tantawi et al. [34, 35, 39] used a 3-μm thick porous silicon membrane that separated two chambers to characterize the epithelial sodium channel (ENAC) proteins.

4 MEMS Actuators

MEMS actuators can be classified by the principle of operation on which they perform work. Actuators may be based on electrostatic attracting and repulsive forces, or may be based on piezoelectric properties, or thermal expansion, magnetic, or fluid pressure. An actuator is any device or system that converts input signal into action, i.e., conversion of one form of energy into the other. MEMS actuators have been used for several critical applications such as drug delivery, transmission, linear and rotary micromotors, microgears, printers, microvalves, micropumps, switches, tweezers, and tongs. MEMS devices and systems offer several advantages over macroscale devices and systems: (1) small size and form factor, (2) high precision and accuracy, (3) low cost, (4) quick response, (5) higher resonance frequency, and (6) integration using existing semiconductor fabrication processes for batch production. Figure 10.12 shows several MEMS actuators designed and fabricated using semiconductor-based fabrication techniques.

Figure 10.11a shows an MEMS-based 2DOF nanopositioning stage is developed, which integrates combo-drive actuators and capacitive displacement sensors to allow for simultaneous actuation and position sensing. Figure 10.11b shows Abbas et al. using MEMS actuators used for tensile testing of platinum thin film to study mechanical properties and failure mechanisms. Figure 10.11c shows a ratcheting system that was fabricated in the five-level technology at MEMS scale, i.e., twenty gears fit on a period in a newspaper sentence. Figure 10.11d shows an *MEMS Mirror*: "Hinged Silicon Mirror and Drive Motors" fabricated by Sandia National Laboratory.

4.1 Mechanical MEMS Actuator

4.1.1 Electrostatic Actuators

Energy stored in device is given by

$$W = \frac{1}{2}CV^2$$

where C is parallel plate capacitance and V is applied voltage

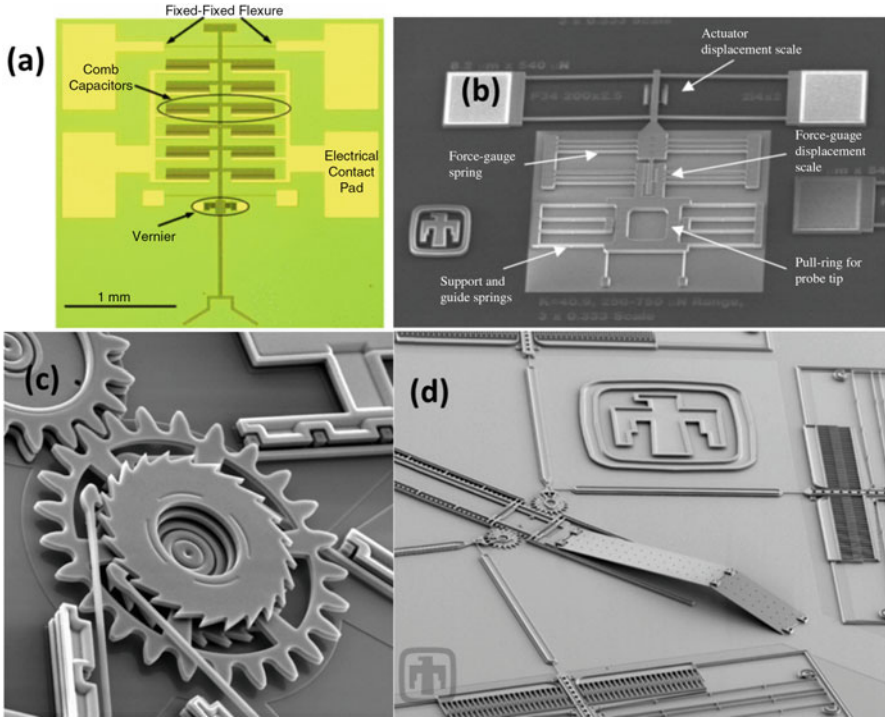The resulting electrostatic force is given by

$$F = \frac{1}{2}V^2 dC/dx$$

Fig. 10.11 Examples of MEMS actuators (**a**) Working principle of a basic MEMS actuator, (Springer itself is the publisher of image) (**b**) SEM image showing force-gauge attached to MEMS actuator with spring combination reprinted with permission from Sandia National Laboratories [41] (**c**) ratcheting system that was fabricated in the five-level technology at MEMS scale (**d**) MEMS Mirror: "Hinged Silicon Mirror and Drive Motors" [40–43]

Fig. 10.12 Working
principle of a comb-drive
MEMS actuator reprinted
from Open access:
Micromachines 2012, 3(2),
396–412;
doi:10.3390/mi3020396 [44]
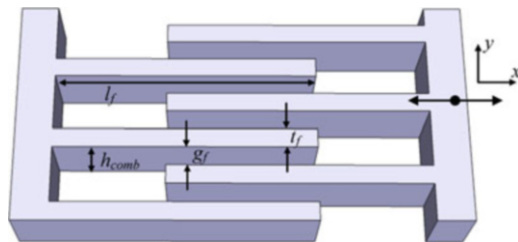
Electrostatic actuator is triggered by forces of electrostatic attraction between two oppositely charged plates. A close parallel plate capacitor can be fabricated using standard MEMS fabrication technique. One example of electrostatic actuator is a combo-drive actuator. The capacitance and the electrostatic force of actuation depend on electrode thickness, topology, and geometry of the device generating 3-dimensional possible translational motion. Figure 10.12 shows the operating principle of a combo-drive MEMS actuator.
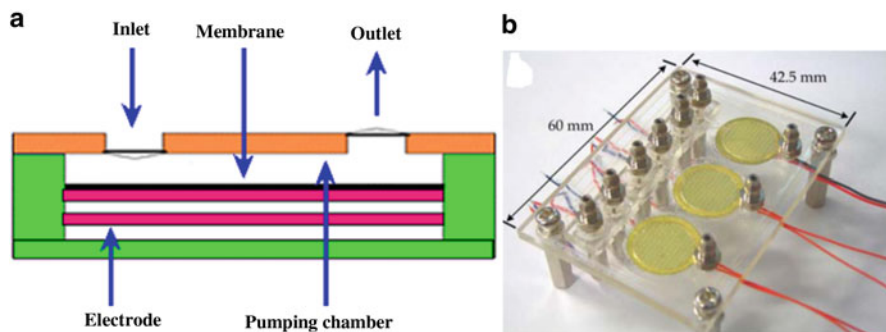
Fig. 10.13 (**a**) Electrostatic micropump repreinted from open access: *Int. J. Mol. Sci.* 2011, *12*(6), 3648–3704 [45] (**b**) Microgear design using birefringence repreinted from open access: *Micromachines* 2012, 3, 279–294 [46]

4.2 Hydraulic/Microfluidic MEMS Actuators

Microfluidic actuators have been used for several applications such as drug delivery, DNA sequencing, nucleic acid analysis, and biochemical analysis. The primary driving factor being miniaturization, these devices provided great accuracy and precision in medical technologies specifically, delivering biological solutions (drugs). The advances in biomedical MEMS technologies lead to possible automation of these important procedures and provided a platform for reduced human requirement, low cost, high sensitivity, high efficiency and efficacy. When choosing materials for biomedical applications, biocompatibility will be the primary factor in the criteria since these devices will be placed inside a human body—just under the skin or somewhere deep inside. Figure 10.13a, b show examples of MEMS micropumps [45, 46].

4.3 Thermal MEMS Actuators: Shape Memory Alloys

Shape memory alloys (SMA) have revolutionized the actuator application, where required blocking force is small and exhibit superelasticity. The material can be bent or shaped at room temperature and returns back to its original shape and size upon heating above the transition temperature thus "actuating" in its return cycle. Figure 10.14 shows the working principle of SMA actuator. Predominantly, Nickel-Titanium alloys exhibit shape-memory effect; however, several other combinations have been studied and demonstrated such as Copper-Zinc-Aluminum, Titanium-Niobium, Copper-Aluminum-Nickel, and Iron-Manganese-Silicon. Shape-memory alloys have been used in several applications. For example—Villanueva et al. used bioinspired shape memory alloy composite actuator for attaining large deformation profiles which was based on rowing mechanism of jellyfish for locomotive motion [47].
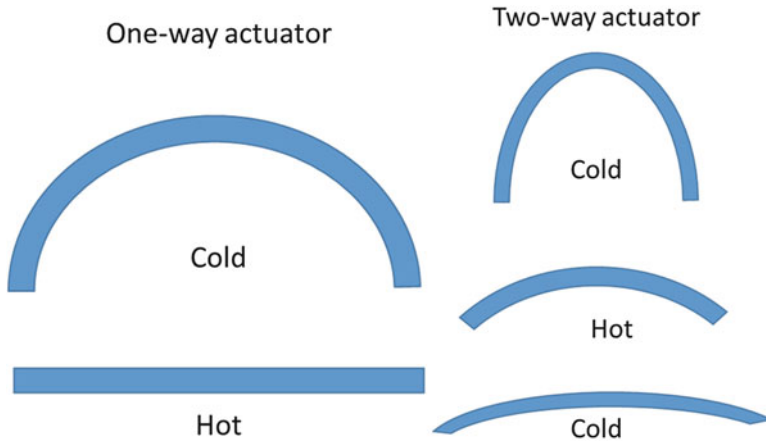
Fig. 10.14 Working principle of one-way and two-way SMA actuators

4.4 Piezoelectric MEMS Actuators

Piezoelectric sensors use direct or converse "piezoelectric effect." A direct piezoelectric effect can be defined as induction of charge due to application of mechanical stress whereas a converse piezoelectric effect can be defined as generation of stress due to applied electric field. Piezoelectric transducers provide the interface between the mechanical and electrical systems with an efficiency of around 50 % (the conversion from mechanical stress to electric field is proportional to square of the electromechanical coupling factor, k^2 and the magnitude of k_{33} for commonly used piezoelectric ceramics is of the order of 0.7). It has been shown that thin-film unimorph type Lead Zirconate Titanate (PZT)-based transducer can supply 20 nW/mm^2 (3 nW with 150 mm^2 area) when external mechanical vibrations are imposed upon the PZT. Recently, there has been success in synthesizing high performance single crystals with better coupling coefficient (k) and energy density. However, implementation of the single crystal piezoelectrics is challenging on the micro/nano platform. Textured thin films of these compositions can be grown, but the coefficients are compromised and the device fabrication requires multiple steps. At this juncture, the best power density reported for the piezoelectric energy harvesting devices is in the range of 0.3–0.5 μW/mm^3. Further improvements in this magnitude requires discovery of materials and mechanisms. Piezoelectric polycrystalline materials are designed specifically to enhance the material charge and voltage constants thereby improving the energy harvested under ambient sources such as mechanical vibrations. Figure 10.15 shows few examples of piezoelectric MEMS actuators. Figure 10.15a shows an working principle of piezoelectric bimorph MEMS actuator [48]. The applied voltage generates deflection of the beam thus, actuating the device. The displacement of the beam is highest at the tip and is dependent on applied voltage. Figure 10.15b shows a piezoelectric micropump
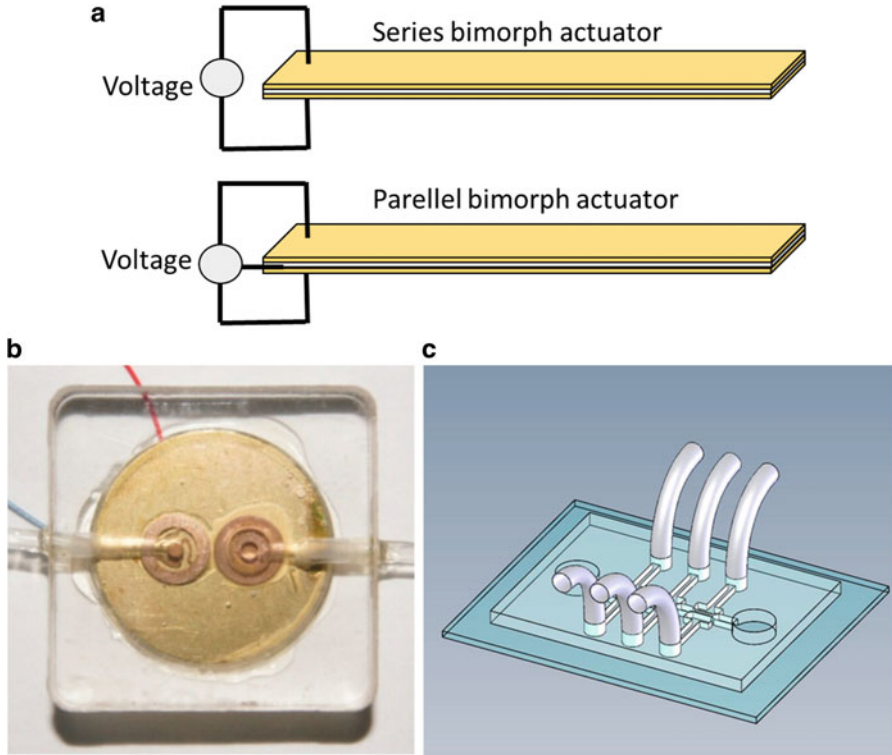
Fig. 10.15 (**a**) Working principle of piezoelectric bimorph MEMS actuator (**b**) Piezoelectric micropump reprinted from open access: *Micromachines* 2015, *6*, 1745–1754 [49] (c) Micropump using three-pair moving structures in double-side mode reprinted from open access: Micromachines 2015, *6*, 216–229 [50]

using a disc vibrator for flow control with low power consumption. Figure 10.15c shows a double side mode micropump using three pairs of moving structures.

Piezoelectric MEMS devices can provide large displacement actuators and ultra-high sensitive sensors operating at wide range of voltages from several millivolts to several tens of volts. On the other hand, in traditional MEMS devices a high voltage and small electrode gap spacing are required to charge a high voltage pump.

PZT can be tuned based on stoichiometry in order to enhance material properties to meet requirements of an MEMS actuator such as high dielectric constant, high piezoelectric coefficient, high mechanical quality factor, high coupling coefficient, and pyroelectric effect. These advantages of PZT make it material of interest in many commercial applications in order to integrate thin films of PZT into MEMS devices and systems [50]. Sano et al. fabricated a multilayer PZT thin film by sputtering for MEMS actuator application. Tip displacement and measured blocking force were much higher than for unimorph and bimorph configurations [51].
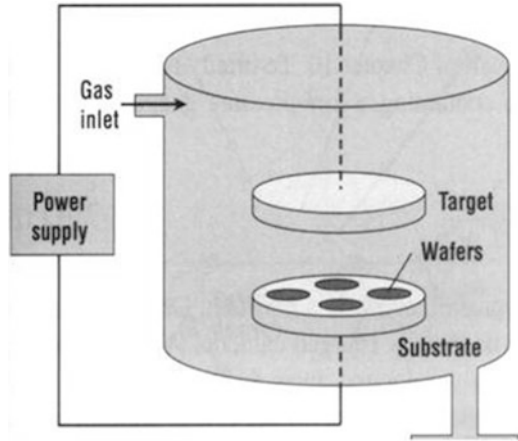
4.5 Magnetic MEMS Actuators

Magnetic actuation is based on the principle of Faraday's law of electromagnetic induction. If one conductor comes in the vicinity of another conductor, then the magnetic fields generated by the two conductors interact thus generating electromotive force.

The Lorentz force generated is given by

$$F = qV \times B$$

where F is the Lorentz force, q is the charge, V is the velocity, and B is the intensity of magnetic field. The force can cause a mechanical motion of the structure. The most common example of magnetic actuation is CD, DVD, and blu-ray disc writing. As the technology moved towards shorter wavelength of laser excitation, the ability to read and write more in a given area improved. For example, a compact disc (CD) could store only 700 MB whereas a DVD disc could hold up to 4.7 GB of data and Blu-ray holds up to 25 GB on single layer. This is due to the reduction of wavelength from 780 nm to 685 nm to 405 nm, respectively. Some materials of choice for MEMS actuators are permanent magnets embedded in polymeric matrix and often chosen based on ease of fabrication and machinability. The magnetic material is integrated into an MEMS device using standard fabrication techniques such as sputtering, electroplating, screen printing, and pulsed laser deposition so that accuracy and precision can be maintained for MEMS device. A typical sputtering setup is shown in Fig. 10.16.

4.6   Magnetostrictive MEMS Actuators

Magnetostriction is the property of a material to change shape under changing magnetic field. The highest coercive filed material found so far is Terfenol-D whereas the highest magnetic permeability is found in Metglas. Nickel can also be used as it has very low coercive field and switching of domains requires less energy.

4.7   Electronics/RF MEMS Actuators

As explained earlier, Aluminum Nitride (AlN) has been used in MEMS actuators. Sandia National Lab designed and fabricated a microresonator RF MEMS based on Aluminum Nitride to provide ultra-miniature RF filters and oscillators with a wide range of frequency and high performance. This technology provides a great solution for radios because they provide a wide frequency range of 32 kHz to 10 GHz. The AlN microresonators were fabricated on a single IC chip, and they can be monolithically integrated with transistors and electronics such as switches and amplifiers circuitry [53]. The single chip microresonator provides microscale center frequency filters and oscillators that are programmable and frequency range of HF to X-band range. Applications include wide ranges of communication systems.

Figure 10.17 shows a 16-channel Aluminum Nitride Filter Array with 1.3 MHz bandwidth, conversion of HF band to X band, fast spectrum analysis, anti-jam, and secure adaptive RF front ends and usage for multiband radios developed by Sandia National Lab [53].
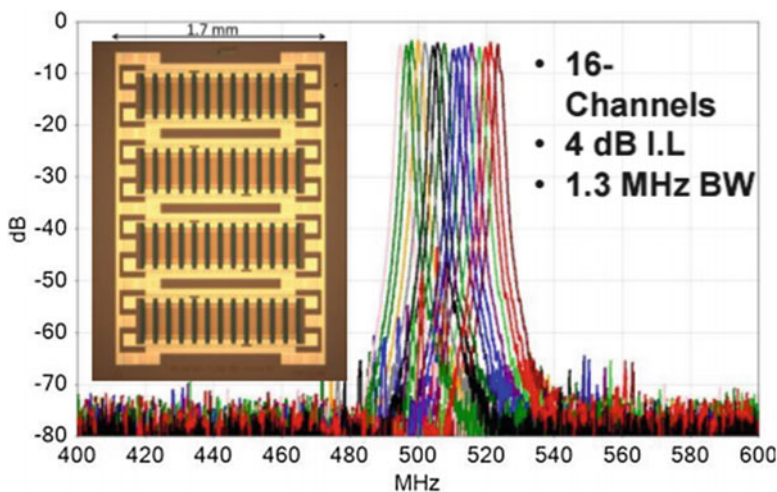

Fig. 10.17 A 16-channel Aluminum Nitride Filter Array [53] reprinted with permission from Sandia National Laboratories

4.8 Summary

In this chapter, a detailed overview of MEMS-based sensors and actuators was presented. A wide range of materials are used in MEMS applications such as metals and alloys, ceramics, polymers, silicon, and glasses. MEMS device fabrication involves steps such as sputtering, deposition, etching, patterning, and several other microfabrication processes. The application often drives the design and materials used for MEMS devices. Due to their low footprint, low cost of fabrication, better material properties, better enhanced performance; MEMS devices are extremely useful in several fields such as medical fields—surgery and data logging, transmission, data analysis, drug delivery, microfluidics—micropumps, gas sensors, mechanical actuators, and so on. The future of MEMS devices is perhaps enhancing this technology and further reducing the footprint of these devices for improved sensitivity and actuation using NEMS—Nano Electro Mechanical Systems. Carbon-based nanomaterials provide a wide variety of properties and form factors that can be realized in commercial applications.

References

1. Madou MJ (2002) Fundamentals of microfabrication: the science of miniaturization, 2nd edn. CRC, Boca Raton, pp 187–189
2. Boustany M (2014) Fingerprint Sensors for Handsets and Tablets—2014, IHS Technology, Inc MEMS & Sensors for Consumer and Mobile forecast report.
3. Dixon R (2015) Powertrain Sensor Market Tracker, IHS Technology, Inc MEMS & Sensors forecast report
4. Bouchaud J (2014) Bosch is Top MEMS Maker in 2013 Top 20 Combined Revenue Reaches $7 Billion; Sensors; March 14, 2014, page 8
5. Cmglee (2015) Wikimedia commons. 23 Mar 2015. [Online]. https://upload.wikimedia.org/wikipedia/commons/2/23/Cmglee_Cambridge_Science_Festival_2015_da_Vinci.jpg. Accessed 26 Jan 2016
6. West J, Halas N (2000) Applications of nanotechnology to biotechnology: commentary. Curr Opin Biotechnol 11:215–217
7. LaVan D, McGuire T, Langer R (2003) Small-scale systems for in vivo drug delivery. Nat Biotechnol 21:1184–1191
8. Lin V, Motesharei K, Dancil KP, Sailor M, Ghadiri MR (1997) A porous silicon-based optical interferometric biosensor. Science 278:840–843
9. Vaseashta A, Dimova-Malinovska D (2005) Nanostructured and nanoscale devices, sensors and detectors. Sci Technol Adv Mater 6:312–318
10. Stewart MP, Buriak JM (2000) Chemical and biological applications of porous silicon technology. Adv Mater 12:859–869
11. Martin-Palma R, Manso-Silvan M, Torres-Costa V (2010) Biomedical applications of nanostructured porous silicon: a review. J Nanophoton 4:042502
12. Park J, Gu L, Maltzahn G, Ruoslahti E, Bhatia S, Sailor M (2009) Biodegradable luminescent porous silicon nanoparticles for in vivo applications. Nat Mater 8:331–336
13. Dietrich T, Ehrfeld W, Lacher M, Krämer M, Speit B (1996) Fabrication technologies for microsystems utilizing photoetchable glass. Microelectron Eng 30:497–504

14. Tantawi KH, Oates J, Kamali-Sarvestani R, Bergquist N, Williams JD (2011) Processing of photosensitive APEX™ glass structures with smooth and transparent sidewalls. J Micromech Microeng 21:017001
15. Williams J, Schmidt C, Serkland D (2010) Processing advances in transparent Foturan glass MEMS. Appl Phys A 99:777–782
16. Cheng Y, Tsai HL, Sugioka K, Midorikawa K (2006) Fabrication of 3D microoptical lenses in photosensitive glass using femtosecond laser micromachining. Appl Phys A 85:11–14
17. Tantawi K, Gaillard W, Helton J, Waddell E, Mirov S, Fedorov V, Williams J (2013) In-plane spectroscopy of microfluidic systems made in photosensitive glass. Microsyst Technol 19(2):173–177
18. Bhardwaj V, Simova E, Corkum P, Rayner D, Hnatovsky C, Taylor R, Schreder B, Kluge M, Zimmer J (2005) Femtosecond laser-induced refractive index modification in multicomponent glasses. J Appl Phys 97:083102
19. Livingston F, Adams P, Helvajian H (2005) Influence of cerium on the pulsed UV nanosecond laser processing of photostructurable glass ceramic materials. Appl Surf Sci 247:526–536
20. Livingston F, Helvajian H (2005) Variable UV laser exposure processing of photosensitive glass-ceramics: maskless micro- to meso-scale structure fabrication. Appl Phys A 81:1569–1581
21. Tantawi K, Waddel E, Williams J (2013) Structural and composition analysis of Apex™ and Foturan™ photodefinable glasses. J Mater Sci 48:5316–5323
22. Cai D, Neyer A, Kuckuk R, Heise H (2008) Optical absorption in transparent PDMS materials applied for multimode waveguides fabrication. Opt Mater 30:1157–1161
23. Mamanee W, Tuantranont A, Afzulpurkar NV, Porntheerapat N, Rahong S, Wisitsoraat A (2006) PDMS based thermopneumatic peristaltic micropump for microfluidic systems. J Phys Conf Ser 34:564–569
24. Wu J, Fedder G, Carley L (2004) A low-noise low-offset capacitive sensing amplifier for a 50-$\mu g/\sqrt{Hz}$ monolithic CMOS MEMS accelerometer. IEEE J Solid-State Circuits 39(5):722–730
25. Wang Y, Yeow JTW (2009) A review of carbon nanotubes-based gas sensors. J Sens 2009. Article ID 493904, 24 pages
26. Bogue R (2007) MEMS sensors: past, present and future. Sens Rev 27(1):7–13
27. Kotovksy J, Tooker A, Horsley D (2010) Thin silicon MEMS contact-stress sensor. In: Hilton Head solid-state sensors, actuators, and microsystems workshop, Hilton Head, SC
28. Park J, Kim M, Lee Y, Lee HS, Ko H (2015) Fingertip skin-inspired microstructured ferroelectric skins discriminate static/dynamic pressure and temperature stimuli. Sci Adv 1(9):e1500661
29. Poor A (2015) Health tech insider. Center Ring Media, 14 Dec 2015. [Online]. http://healthtechinsider.com/tag/piezoelectric/. Accessed 20 Jan 2016
30. Dirksen JA, Duval K, Ring TA (2001) NiO thin-film formaldehyde gas sensor. Sens Actuat 80:106–115
31. Karunagaran B, Uthirakumar P, Velumani SCS, Suh E-K (2007) TiO_2 thin film gas sensor for monitoring ammonia. Mater Charact 58:680–684
32. Li J, Lu Y, Ye Q, Cinke M, Han J, Meyyappan M (2003) Carbon nanotube sensors for gas and organic vapor detection. Nano Lett 3:929–933
33. Bronikowski MJ, Willis PA, Colbert DT, Smith KA, Smalley RE (2001) Gas-phase production of carbon single-walled nanotubes from carbon monoxide via the HiPco process: a parametric study. J Vac Sci 19:1800–1805
34. Tantawi KH, Berdiev B, Cerro R, Williams JD (2013) Porous silicon membrane structure for investigation of transmembrane proteins. J Superlattices Microstruct 58:72–80
35. Tantawi KH (2013) Porous silicon platform for investigation of transmembrane proteins: an artificial cell membrane. Lambert Academic Publishing, Engelska
36. Buchholz K, Tinazli A, Kleefen A, Dorfner D, Pedone D, Rant U, Tampe R, Abstreiter G, Tornow M (2008) Silicon-on-insulator based nanopore cavity arrays for lipid membrane investigation. Nanotechnology 19:445305–445311

37. Simion M, Ruta L, Mihailiscu C, Kleps I, Bragaru A, Miu M, Ignat T, Baciu I (2008) Porous silicon used as support for protein microarray. Superlattices Microstruct 11:1–8
38. Worsfold O, Voelcker NH, Nishiya T (2006) Biosensing using lipid bilayers suspended on porous silicon. Langmuir 22:7078–7083
39. Tantawi KH, Cerro R, Berdiev B, Martin MED, Montes FJ, Patel D, Williams JD (2013) Investigation of transmembrane protein fused in lipid bilayer membranes supported on porous silicon. J Med Eng Technol 37:28–34
40. Mousavi K, Leseman A, Chad Z (2012) Basic MEMS Actuators. In: Bhushan B (ed) Encyclopedia of Nanotech-nology. Springer, Dordrecht, pp 173–185. doi:10.1007/978-90-481-9751-4_292. ISBN 978-90-481-9751-4
41. Michael S. Baker et al. Final Report: Compliant Thermo-Mechanical MEMS, Actuators LDRD #52553. Sandia National Labs SAND2004-6635.
42. Sandia National Labs http://www.sandia.gov/media/NewsRel/NR1999/layer.htm.
43. http://ccar.colorado.edu/asen5050/projects/projects_2001/eby/mems.html.
44. Groen MS, Brouwer DM, Wiegerink RJ, Lötters JC (2012) Design Considerations for a Micro-machined Proportional Control Valve. Micromachines 3(2):396–412. doi:10.3390/mi3020396
45. Ashraf MW, Tayyaba S, Afzulpurkar Int N (2011) J Mol Sci 12(6):3648–3704
46. Lam RHW, Li WJ (2012) A Digitally Controllable Polymer-Based Microfluidic Mixing Module Array. Micromachines 3:279–294
47. Villanueva AA, Joshi KB, Blottman JB, and Priya S (2010) A bio-inspired shape memory alloy composite (BISMAC) actuator. Smart Materials and Struc-tures 19(2)
48. Bedekar V, Oliver J, Priya S (2010) Design and fabrication of bimorph trans-ducer for optimal vibration energy harvesting. IEEE Transactions on Ultrason-ics, Ferroelectrics, and Frequency Control 57(7):1513–1523. doi:10.1109/TUFFC.2010.1582
49. Chen S, Liu Y, Shen Y, Wang J, Yang Z (2015) The structure of wheel check valve influence on air block phenomenon of piezoelectric micro-pump. Micromachines 6:1745–1754
50. Chiou C-H, Yeh T-Y, Lin J-L (2015) Deformation analysis of a pneumatical-ly-activated poly-dimethylsiloxane (PDMS) membrane and potential micro-pump applications. Micromachines 6:216–229
51. Sano R, Inoue J, Kanda K, Fujita T, Maenaka K (2015) Jpn J Appl Phys 54(10S). http://iopscience.iop.org/article/10.7567/JJAP.54.10ND03/meta)
52. Wikimedia Commons https://en.wikipedia.org/wiki/MEMS_magnetic_actuator#/media/File:Sputtering1.jpg.
53. Microresonators for Advanced RF Devices, Roy H. Olsson III, Ken Wojcie-chowski and Chris Nordquist, Sandia National Lab, 2013, 207–224

Chapter 11
Soot Load Sensing in a Diesel Particulate Filter Based on Electrical Capacitance Tomography

Ragibul Huq and Sohel Anwar

Abstract This work presents a novel approach to particulate material (soot) measurement in a diesel particulate filter (DPF) using electrical capacitance tomography (ECT). Modern diesel engines are equipped with DPFs, as well as onboard technologies to evaluate the status of DPF because complete knowledge of DPF soot loading is very critical for robust and efficient operation of the engine exhaust after treatment system. Emission regulations imposed upon all internal combustion engines including diesel engines on gaseous as well as particulate (soot) emissions by environment regulatory agencies. In course of time, soot will be deposited inside the DPFs which tend to clog the filter and hence generate a back pressure in the exhaust system, negatively impacting the fuel efficiency. To remove the soot buildup, regeneration of the DPF must be done as an engine exhaust after treatment process at predetermined time intervals. Passive regeneration increases the exhaust heat to burn the deposited soot while active regeneration injects external energy in, such as injection of diesel into an upstream diesel oxidation catalyst (DOC), to burn the soot. Since the regeneration process consumes fuel, a robust and efficient operation based on accurate knowledge of the particulate matter deposit (or soot load) becomes essential in order to keep the fuel consumption at a minimum. Here we propose a sensing method for a DPF that can accurately measure in-situ soot load using ECT. Lab experimental results show that the proposed method offers an effective way to accurately estimate the soot load in DPF. The proposed method is expected to have a profound impact in improving overall DPF efficiency (and thereby fuel efficiency), and durability of a DPF through appropriate closed loop regeneration operation.

R. Huq
Cummins, Inc., Columbus, IN, USA
e-mail: Ragibul.Huq@cummins.com

S. Anwar (✉)
Department of Mechanical Engineering, Purdue School of Engineering and Technology, Indiana Univ Purdue Univ Indianapolis, 723 W. Michigan Street, SL 260F, Indianapolis, IN 46202, USA
e-mail: soanwar@iupui.edu

Nomenclature

A	Electrode surface area
C	Normalized electrode-pair capacitances
C_m	Overall capacitance in Maxwell Garnett Permittivity Model
C_p	Overall capacitance in Parallel Permittivity Model
C_s	Overall capacitance in Series Permittivity Model
d	Distance between two plates
D	Flux density
E	Electric field strength between the plates
f_i	Volume fraction occupied by the inclusions of the i-th sort
K	Normalized pixel permittivity's matrix
N_{ik}	Depolarization factors of the i-th sort of inclusions
Q	Charge
S	Sensitivity matrix
V	Potential difference
α	Temperature coefficient
ε	Permittivity
ε_b	Relative permittivity of a base dielectric
ε_i	Relative permittivity of the i-th sort of inclusions
ε_m	Effective permittivity in Maxwell Garnett Permittivity Model
ε_o	In vacuum, the value of $\varepsilon_o = 8.854 \times 10^{-12}$ F/m
ε_p	Effective permittivity in Parallel Permittivity Model
ε_r	Relative permittivity
ε_s	Effective permittivity in Series Permittivity Model
ρ_T	Resistivity at temperature T

1 Introduction

Particulate matter (PM) also known as particle pollution or soot is a complex
mixture of extremely small particles and liquid droplets. Particle pollution is made
up of a number of components, including acids (such as nitrates and sulfates),
organic chemicals, metals, and soil or dust particles. The size of particles is directly
linked to their potential for causing health problems. US Environment Protection
Agency (US EPA) is concerned about particles that are 10 μm in diameter or smaller
because those are the particles that generally pass through the throat and nose and
enter the lungs according to EPA website, and once inhaled, these particles can
affect the heart and lungs and cause serious health effects [1].

Fig. 11.1 Porous walls of a DPF [3]
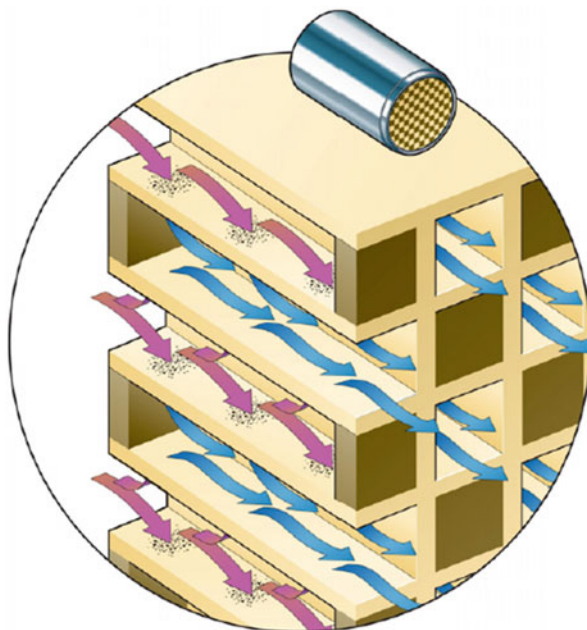
Diesel particulate filters (DPFs) were first used in 1980 which remove the PM/soot from the exhaust of the diesel engine with an efficiency level of 90 % or more. The most commonly used DPFs are porous ceramic wall-flow filters, as shown schematically in Fig. 11.1. Refractory materials such as Silicon Carbide, Cordierite, or Aluminum-Titanate are used for this purpose [2]. Silicon carbide filters dominate the market owing to the material's mechanical strength and high thermal stability [3]. Alternate channels are plugged, forcing the exhaust through the porous channel walls. The gaseous exhaust passes through the porous walls, but PM is trapped in the filter.

However, as the PM or soot is retained by the filter, the filter passageway increasingly becomes more restrictive resulting in elevated backpressure in the exhaust. This furthers results in lower fuel efficiency for the engines since the pistons have to exert more pressure to purge the exhaust gas. One effective way to address this problem is to burn the soot load in the DPF periodically either by injecting more fuel in the engine or by a separate combustor upstream of the DPF with the aid of a diesel oxidation catalyst (DOC). The latter which is known as active regeneration of DPF is more efficient and is commonly used for DPF. Here a fuel doser is used to raise the exhaust gas temperature to burn off the soot load in DPF. The timing and amount of fuel dosing is critical in ensuring optimal performance of DPF functions. Current commercially available DPF PM detection methods are mostly based on pressure differential.

The performance efficiency of a DPF with active regeneration is largely dependent on the accuracy of soot load estimation. Current soot load estimation is based

on differential pressure measurement across the DPF whose accuracy can vary up to ±50 % from the true soot load [4, 5]. As a result, fuel dosing for active regeneration may not be optimal. It has been shown that fuel penalty caused by regeneration (2.2–5.3 %) is more than fuel penalty due to backpressure (1.5–2.0 %) [6]. Accurate soot load knowledge is also necessary if one wants to rule out possible overheating of the DPF caused by exothermic soot oxidation. Because if DPF is allowed to accumulate too much PM, the large amount of heat released upon regeneration cannot effectively be dissipated, resulting filter damage such as by the formation of cracks or regions which may be locally melted.

Rose and Boger [7] describe soot mass estimation using pressure drop signal combined with other parameters. This team basically focuses on the real-world accuracy during highly transient operation. The geometric intricacies also affect the pressure drop. Differential pressure detection has few limitations so researchers are pursuing other means for detecting the soot deposition. Husted et al. [8] in their paper "sensing of particulate matter for on-board diagnosis of particulate filters" explore the possibilities of soot measurement using particulate material sensor placed at the exhaust stream. With the deposition of the soot along the surface of the sensor the electrical conductivity will increase and is measured as part of an electrical circuit. This sensing method required robust design due to its placement directly inside the exhaust stream with exposure to urea, water impingement, poison, ash, acid, and very high temperature (e.g., 900 °C).

Fischerauer et al. used microwave technology to measure the soot load inside DPF [3]. This approach is based on the interaction of electromagnetic fields or waves at microwave frequencies with the DPF. Researchers assumed that the addition of soot in a dielectric medium would influence its permittivity and conductivity, observable via resonance frequency shift or via changes in the attenuation of the resonance peaks (cavity perturbation). The main concern regarding this measuring method is about cost and robustness. Sappok et al. [4] developed radio frequency based soot load sensor. Two antennas are installed on either side of a DPF that transmit and receive the radio frequency signals. Signal amplitude and transmitted power is a function of the losses within the material through which the wave travels. With the increase of DPF soot deposition, dielectric properties also change which affect the RF signal directly. Although the RF sensor response showed good results in both steady state and transient condition, its long-term reliability is yet to be established. It is also very intrusive since the antennas must be inserted in the exhaust flow path causing some level of flow resistance (negatively affecting fuel efficiency). GE introduced Accusolve DPF soot load sensor based on RF based technology.

Strzelec et al. [9] from Oak Ridge national laboratory proposed nondestructive neutron computed tomography measurement of DPF to measure soot. Basic principle of this method is whenever a neutron beam is passing through matter then there will an attenuation of the beam caused by absorption and scattering of material. By measuring the intensity of the attenuated signal, material properties can be detected.

Using this method quantification of soot loading can be done but using the whole neutron beam setup for real-time soot measurement in an automotive engine or a power generator engine is still very expensive.

It is clear that a sensor for soot load measurement must be accurate, robust, and cost effective. In this chapter, we present a soot load sensing that tries to achieve these desired characteristics. This work involves the development of a soot load sensor using electrical capacitance tomography (ECT) which is shown to be more accurate, robust, and inexpensive. This sensor can be used in the feedback loop to improve the soot load estimation allowing for correct amount of fuel injection upstream of DOC and thus potentially improving the overall DPF performance.

2 Modeling of Electrical Capacitance Tomography

2.1 Permittivity Model

Modeling of a complete ECT DPF soot measurement system requires three subsystems: ECT sensor, DPF model, and data acquisition system. The ECT sensor has two parallel conducting plates separated by an air gap which are connected through a switch and a resistor to a signal generator. When the switch is closed, electrons are drawn from the upper plate through the resistor to the positive terminal of the signal generator. This action creates a net positive charge on the top plate. Electrons are being repelled by the negative terminal through the lower conductor to the bottom plate at the same rate they are being drawn to the positive terminal. This transfer of electrons continues until the potential difference across the parallel plates is exactly equal to the battery voltage.

If a potential difference of V volts is applied across two plates separated by a distance of d, the electric field strength between the plates is determined by E.

$$E = {V}/{d} \tag{11.1}$$

The ratio of the flux density to the electric field intensity in the dielectric is called the permittivity of the dielectric [10].

$$\varepsilon = {D}/{E} \tag{11.2}$$

For a vacuum, the value of ε (denoted by ε_o) is $\varepsilon_o = 8.854 \times 10^{-12}$ F/m. The ratio of the permittivity of any dielectric to that of a vacuum is called the relative permittivity, ε_r. It simply compares the permittivity of the dielectric to that of air.

$$\varepsilon_r = {\varepsilon}/{\varepsilon_o} \tag{11.3}$$

If the charge Q is large enough, then flux density D can be written as:

$$D = \frac{Q}{A}$$

$$\varepsilon = \frac{D}{E} = \frac{Q/A}{V/d} = \frac{Qd}{VA}$$

Again $C = \frac{Q}{V}$.
Therefore, $\varepsilon = Cd/A$.

$$C = \varepsilon A/d$$
$$\text{Or,} \quad C = \varepsilon_0 \varepsilon_r \frac{A}{d} \tag{11.4}$$

Based on the abovementioned principle and Eq. (11.4), ECT system has been established. The capacitance measured depends on the relative permittivity of the materials between the electrodes. Their relation can be linear or nonlinear in material type. It also depends on the permittivity models used to characterize the way in which the material fills the air gap space. The various permittivity models are used, namely the series model, parallel model, and Maxwell's model.

Let's assume that one substance has a relative permittivity ε_r in the air gap. If that substance occupies x volume of the total space between two electrodes, the effective permittivity of the mixture and their dependence on capacitance can be computed as described in the following sections.

2.1.1 Series Permittivity Model

Assuming that two materials with relative permittivity ε_r and ε_o in a pipe lie on top of one another, the effective capacitance can be considered as two capacitances connected in series. This is illustrated in Fig. 11.2. In this case, the capacitance and
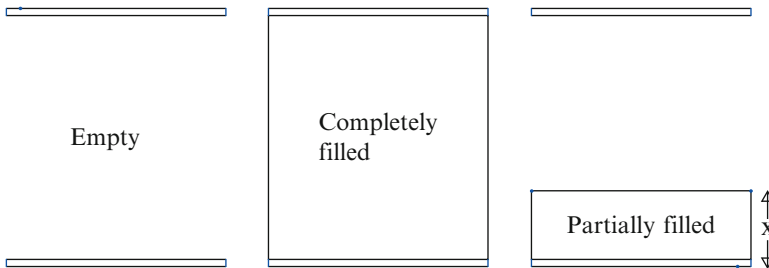

Fig. 11.2 Empty, completely filled, and partially filled
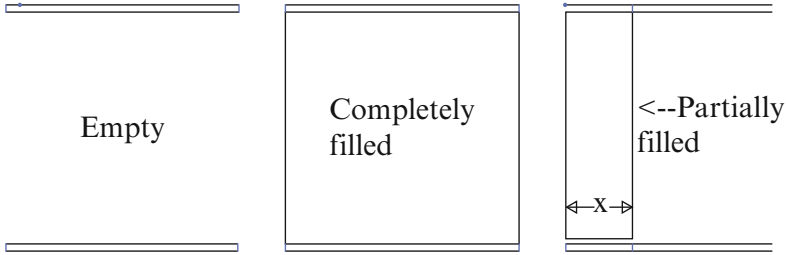
Fig. 11.3 Empty, completely filled, and partially filled

permittivity are related in a nonlinear fashion. The effective permittivity and overall capacitance are given respectively in Eqs. (11.5) and (11.6) [11].

$$\varepsilon_s = \frac{\varepsilon_r \varepsilon_o x \left(1 - x\right)}{1 - x \left(\varepsilon_r - 1\right)} \tag{11.5}$$

$$C_s = {A\varepsilon_s}\big/{d} \tag{11.6}$$

2.1.2 Parallel Permittivity Model

If the materials with relative permittivity ε_r and ε_o in the pipe appear as discrete band and appear side by side, their effective capacitance can be considered as two capacitances connected in parallel. This is illustrated in Fig. 11.3. In this case, the capacitance and permittivity are related linearly. The effective permittivity and overall capacitance are given respectively in Eqs. (11.7) and (11.8) [11].

$$\varepsilon_p = \varepsilon_o \left[1 + x \left(\varepsilon_r - 1\right)\right] \tag{11.7}$$

$$C_p = {A\varepsilon_p}\big/{d} \tag{11.8}$$

2.1.3 Maxwell Garnett Permittivity Model

The generalized Maxwell Garnett permittivity model provides the mixing formula for multiphase mixtures with randomly oriented ellipsoidal inclusions [12]. In this model, two materials of relative permittivity ε_i and ε_b are mixed randomly as in Fig. 11.4, then the effective permittivity becomes ε_m. Here f_i is the volume fraction occupied by the inclusions of the i-th type, and N_{ik} are the depolarization factors of the i-th type of inclusions.
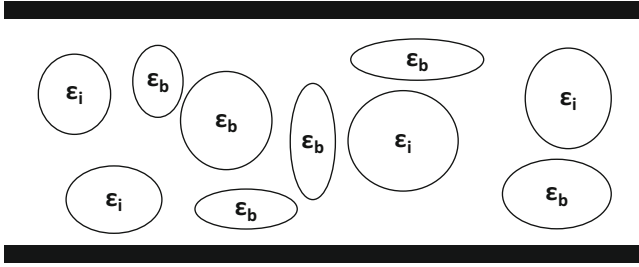
Fig. 11.4 Multiphase mixtures inside capacitor

$$\varepsilon_{\mathrm{m}} = \varepsilon_{\mathrm{b}} + \frac{\frac{1}{3}\sum_{i=1}^{n} f_i\,(\varepsilon_i - \varepsilon_{\mathrm{b}})\sum_{k=3}^{3}\dfrac{\varepsilon_{\mathrm{b}}}{\varepsilon_{\mathrm{b}} + N_{ik}\,(\varepsilon_i - \varepsilon_{\mathrm{b}})}}{1 - \frac{1}{3}\sum_{i=1}^{n} f_i\,(\varepsilon_i - \varepsilon_{\mathrm{b}})\sum_{k=3}^{3}\dfrac{N_{ik}}{\varepsilon_{\mathrm{b}} + N_{ik}\,(\varepsilon_i - \varepsilon_{\mathrm{b}})}} \tag{11.9}$$

$$C_{\mathrm{s}} = {A\varepsilon_{\mathrm{s}}}\big/{d} \tag{11.10}$$

2.2 Electrical Capacitance Tomography Measurement Principle

ECT is used to obtain information about the spatial distribution of a mixture of dielectric materials inside a vessel, by measuring the electrical capacitances between sets of electrodes placed around its periphery and converting these measurements into an image or graph showing the distribution of permittivity [13]. The images are approximate and of relatively low resolution, but they can be generated at relatively high speeds.

ECT can be used with any arbitrary mixture of different nonconducting dielectric materials such as plastics, hydrocarbons, sand, or glass. However, an important application of ECT is viewing and measuring the spatial distribution of a mixture of two different dielectric materials (a two-phase mixture), as in this case, the concentration distribution of the two components over the cross section of the vessel can be obtained from the permittivity distribution.

An ECT system consists of a capacitance sensor, Capacitance Measurement Unit (CMU), and a control computer. For imaging a single vessel type with a fixed cross-section and with a fixed electrode configuration, the measurement circuitry can be integrated into the sensor and the measurement circuits can be connected directly to the sensor electrodes. This simplifies the measurement of interelectrode capacitances and is potentially a good design solution for standardized industrial sensors.
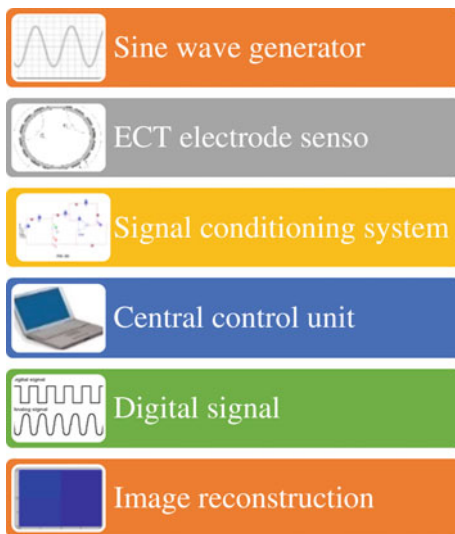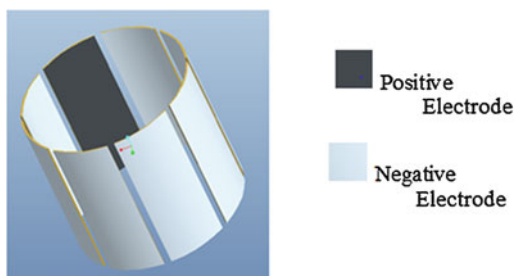
Fig. 11.5 ECT systems flowchart

Fig. 11.6 Capacitance measurement principles

However, most current applications for ECT are in the research sector, where it is preferable to have a standard capacitance measuring unit which can be used with a wide range of sensors. In this case, screened cables connect the sensor to the measurement circuitry, which must be able to measure very small interelectrode capacitances, of the order of 10^{-15} F (1 fF), in the presence of much larger capacitances to earth of the order of 200,000 fF (mainly due to the screened cables) [14]. A diagram of a basic ECT system of this type is shown in Fig. 11.5.

ECT system is a set of capacitor plates placed around a pipe or any other vessel. A source voltage is applied between one electrode (the source electrode) and ground and the resulting currents flowing between the source electrode and the remaining (detector) electrodes to ground are measured. These currents are directly proportional to the capacitances between the source and detector electrodes. The set of capacitance measurements made.

In ECT, a complete set of measurement projections is made by exciting each electrode in turn as a source electrode and measuring the currents which flow into the remaining detector electrodes. So for an 8-electrode sensor, as shown in Fig. 11.6 there will be $8 \times 7 = 56$ possible capacitance measurements. However, as half of

these will be reciprocal measurements (the same capacitance should be measured by exciting electrode 1 as a source and measuring the current into electrode 2 as is obtained by exciting electrode 2 as a source and measuring the current into electrode 1, etc.), there will only be 28 unique capacitance measurements for a complete set of projections. In general for a sensor with E electrodes, there will be $E \times (E-1)/2$ unique capacitance measurements.

The set of measured interelectrode capacitance values and subsequently obtained permittivities are normalized to construct the permittivity images.

$$
\begin{aligned}
\text{Normalized capacitance,} \quad & C_n = \frac{C_i - C_i(\text{emp})}{C_i(\text{full}) - C_i(\text{emp})} \quad 0 < C_n < 1. \\
\text{Normalized permittivity,} \quad & K_n = \frac{K_i - K_i(\text{emp})}{K_i(\text{full}) - K_i(\text{emp})} \quad 0 < K_n < 1.
\end{aligned}
\tag{11.10}
$$

The normalized values are then projected into a square pixel grid where the pixel values are similarly normalized to be between 0 and 1. The image formed is not an exact solution but an approximate solution.

2.3 Tomographic Image Formation

The permittivity image or tomographic images are mapped onto a square pixel grid. The complete set of measured interelectrode capacitance values is required to reconstruct one permittivity distribution image. Figure 11.7 shows a 16×16 square pixel grid used to display the permittivity distribution image of a 4-electrode sensor having circular intersection of DPF.

Fig. 11.7 Square pixel grid
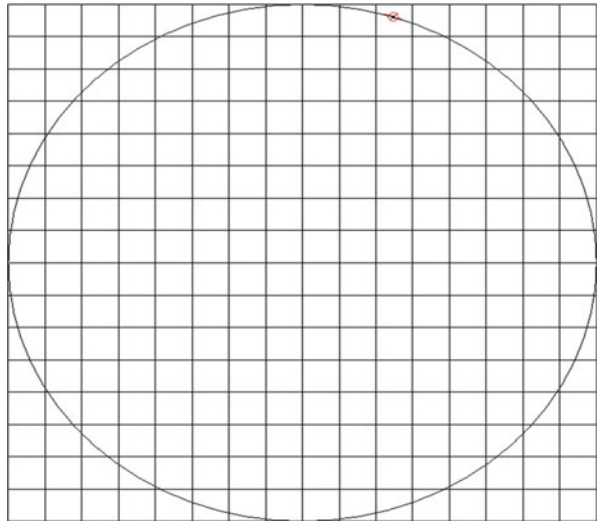
From this (16×16) square pixel grid containing 256 pixels, only 224 are needed to construct the cross-sectional image of the DPF and remaining pixels are not required and hence neglected.

The field lines between two plates are curved and to suit the requirement these lines can be approximated. A proper sensitivity map, purpose of which is to aid in selecting the proper pixel that individually contributes to the capacitance changes, has to be developed for visualizing the electric field established between two electrodes when one of them is excited.

A simple procedure for reconstructing an image of an unknown permittivity distribution inside the sensor from the capacitance measurements is the Linear Back Projection (LBP) algorithm [15]. Although its reconstruction accuracy is not very good, LBP has the advantage of being quite fast, in practice requiring only the multiplication of a fixed reconstruction matrix times the vector of measurements.

Let us consider an x-electrode sensor, from which N number of voltage data has been recorded, and square pixel grid pixel number is M. A gray level $K(M)$ for each pixel has been calculated by the basic LBP formula [16].

$$K_1 = S_{a1} \times C_1. + S_{a2} \times C_2 + S_{a3} \times C_3 + S_{a4} \times C_4 + S_{a5} \times C_5 + \cdots + S_{aN} \times C_N$$
$$K_2 = S_{b1} \times C_1. + S_{b2} \times C_2 + S_{b3} \times C_3 + S_{b4} \times C_4 + S_{b5} \times C_5 + \cdots + S_{bN} \times C_N$$
$$\cdots$$
$$K_M = S_{M1} \times C_1. + S_{M3} \times C_3 + S_{M3} \times C_3 + S_{M4} \times C_4 + S_{M5} \times C_5 + \cdots + S_{MN} \times C_N$$

The relationships presented above between capacitance permittivity distributions can be written in a normalized form as:

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \\ \vdots \\ K_M \end{bmatrix} = \begin{bmatrix} S_{a1} & S_{a2} & S_{a3} & \dots & S_{aN} \\ S_{b1} & S_{b2} & S_{b3} & \dots & S_{bN} \\ S_{c1} & S_{c3} & S_{c3} & \dots & S_{cN} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ S_{M1} & S_{M2} & S_{M3} & \dots & S_{MN} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{bmatrix}$$

$$\mathbf{K = SC} \tag{11.11}$$

$\mathbf{C} = N \times 1$ matrix containing the normalized electrode-pair capacitances C_m (in the nominal range 0–1).

$\mathbf{K} = M \times 1$ matrix containing the normalized pixel permittivities (in the nominal range 0–1) and N is the number of pixels representing the sensor cross-section.

$\mathbf{S} = M \times N$ matrix containing the set of sensitivity matrices for each electrode-pair. This sensitivity map can be defined as

$$S_{Mi} = \frac{C_i - C_i\,(\text{emp})}{C_i\,(\text{full}) - C_i\,(\text{emp})} \quad \text{for } i = 1 \dots N.$$

The sensitivity matrix describes how the measured capacitance between any combinations of electrodes changes when a change is made to the dielectric constant of a single pixel inside the sensor. Here $C_i(emp)$ is the capacitance voltage when DPF is completely empty and $C_i(full)$ is the capacitance voltage when DPF is completely full. The properties of the capacitance sensor are measured or calculated initially to produce a sensor sensitivity matrix for the case when the sensor is empty. Sensitivity matrix is composed of a set of sub-matrices (or maps) whose elements correspond to the individual pixels in a rectangular grid which is used to define the sensor cross-section. The sensor is normally calibrated at each end of the range of permittivity to be measured by filling the sensor with the lower permittivity material initially and measuring all of the individual interelectrode capacitances. This operation is then repeated using the higher permittivity material. The data obtained during the calibration procedure is used to set up the measurement parameters for each measuring channel and is stored in a calibration data file.

In principle, once the set of interelectrode capacitances C has been measured, the permittivity distribution $\mathbf{K}$ can be obtained from these measurements using Eq. (11.11).

Direct contributions of pixels to the measured capacitance between any specific electrode-pair are not be specified, but it can be shown from the sensitivity matrix $\mathbf{S}$ that certain pixels have more effect than others on this capacitance. Consequently, component values allocated to each pixel proportional to the product of the electrode-pair capacitance and the pixel sensitivity coefficient for this electrode-pair. Based on this approximation, the LBP algorithm uses the sensitivity matrix $\mathbf{S}$ which has the dimension $(M \times N)$.

This process is repeated for each electrode-pair capacitance in turn and the component values obtained for each pixel are summed for the complete range of electric pairs.

Previous studies regarding diesel soot dielectric properties have shown that in microwave range dielectric constant of soot has dependency on the soot layer thickness which is captured in Fig. 11.8.

In this work, Series Permittivity Model has been chosen to calculate the effective permittivity of soot and air mixture.

$$\varepsilon_s = \frac{\varepsilon_r \varepsilon_o x (1-x)}{1 - x (\varepsilon_r - 1)}$$

$C_s = {A\varepsilon_s}/{d}$ In the model, the length of DPF is set at 6 in., so the length of the capacitor plate will be 6 in. and for four capacitor plate ECT sensor the width of the capacitor plate will be 4.71 in. (approx.) and maximum distance between two plates will be 5.6 in. So

$$\varepsilon_s = \frac{5 \times 8.854 \times 10^{-12} \times (1-2)}{1 - 2(5-1)} = 1.2649 \times 10^{-11}$$
$$C_s = 6 \times 4.72 \times 1.2649 \times 10^{-11} \big/ 5.6 = 63.8301 \text{ pF} \tag{11.12}$$
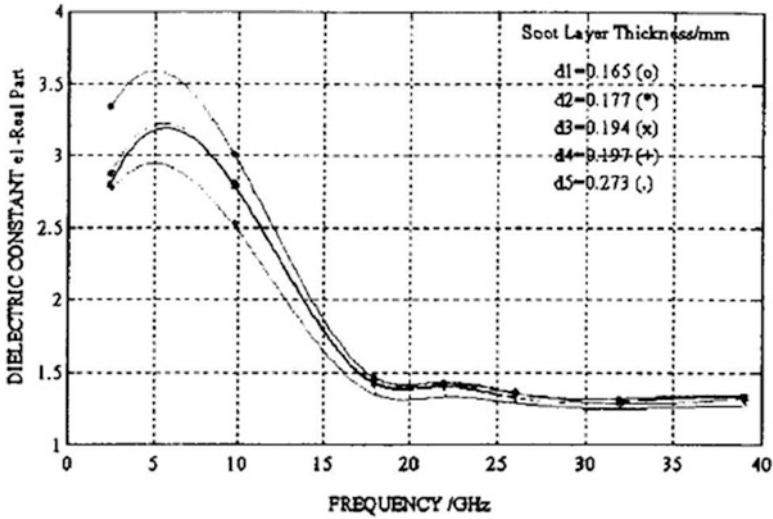
Fig. 11.8 Real part of dielectric constant and soot layer thickness [25]

3   Experimental Setup

The experimentation of soot measuring system using ECT was performed on an experimental model of DPF made of Nylon 66, unreinforced, and flame retardant [17]. This Nylon 66 has very close dielectric properties of conventional DPF made from cordierite which has dielectric constant of 4.7 (approx.) in 1 MHz [18].

The DPF outer shell wall is conductive so the ECT system electrode has to be placed inside of the DPF wall. In that case the components of capacitance due to the electric field inside the sensor will always increase in proportion to the material permittivity when a higher permittivity material is introduced inside the sensor.

The internal temperature of DPF will be highest when the regeneration of soot is taking place. Regeneration is a process of soot removal from the DPF and there are two different approaches existed, one is active and another is passive. Active systems use extra fuel, whether through burning to heat the DPF, or providing extra power to the DPF's electrical system. This process required 600 °C to burn diesel PM. This temperature can be reduced to somewhere in the range of 350–450 °C by use of a fuel borne catalyst [3]. There is a more effective way to burn soot at lower temperature brought by Johnson Matthey's novel two-component design [3]. In this novel approach the catalyst is positioned before the filter to convert NO into NO_2. The NO_2 then oxidizes the soot that is collected on the filter to regenerate the filter at a much lower temperature than is normally required. In fact, the CRT (continuously regenerating technology) enables the filter to be regenerated at a temperature that is 20 % lower than other filters on the market. By using this approach the soot burning temperature can be reduced up to 240 °C. So whatever material we are using as

ECT electrode it has to be able to withstand a versatile range of temperature. The properties that led us to choose copper as ECT electrode are [19]:

1. Melting point at 1357°K or 1084 °C
2. Do not react with water
3. Resistivity $= 1.68 \times 10^{-8}$ (Ωm) at 20 °C
4. Conductivity $= 5.96 \times 10^7$ (s/m) at 20 °C
5. Temperature coefficient $= 0.003862$ (K^{-1})
6. Copper resists corrosion from moisture, humidity, and industrial pollution
7. However products from other carrion like oxide, chloride, and sulfide are conductive.

Due to very low temperature coefficient of copper, the change of conductivity with the change of temperature will be very low. If the assumption made that the temperature change inside of DPF is linear, then relationship between resistivity and temperature will be:

$$\rho(T) = \rho_o \left[1 + \alpha \left(T - T_o\right)\right]$$
$$\rho(T) = 3.489 \times 10^{-8} \, (\Omega m)$$

Figure 11.9 shows an isometric view of a design with eight capacitance electrodes and DPF model of 152 mm (6 in.) length with 130 mm (5 in.) diameter. There is a trade-off in choosing the number of electrodes. Higher number of electrodes means complicated and expensive data acquisition hardware, smaller capacitance to be measured; slower data acquisition as we can see from Table 11.1 and currently 8–12 numbers of electrodes are commonly used in an ECT sensor. These eight
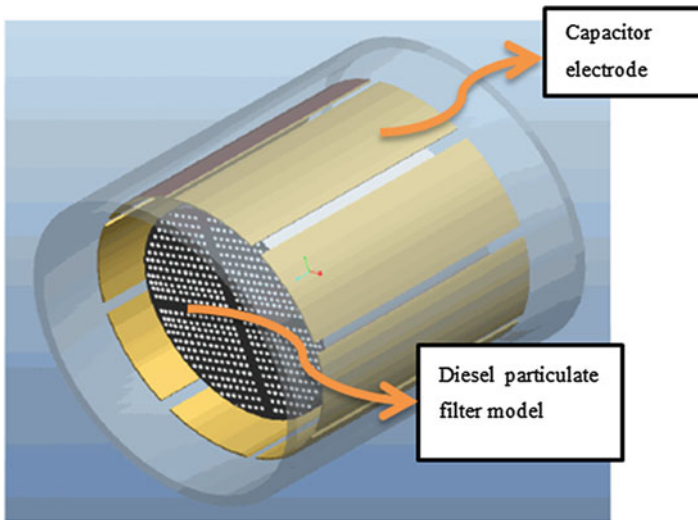

Fig. 11.9 Isometric view of the sensor design

Table 11.1 Number of electrodes vs. number of independent measurement [26]

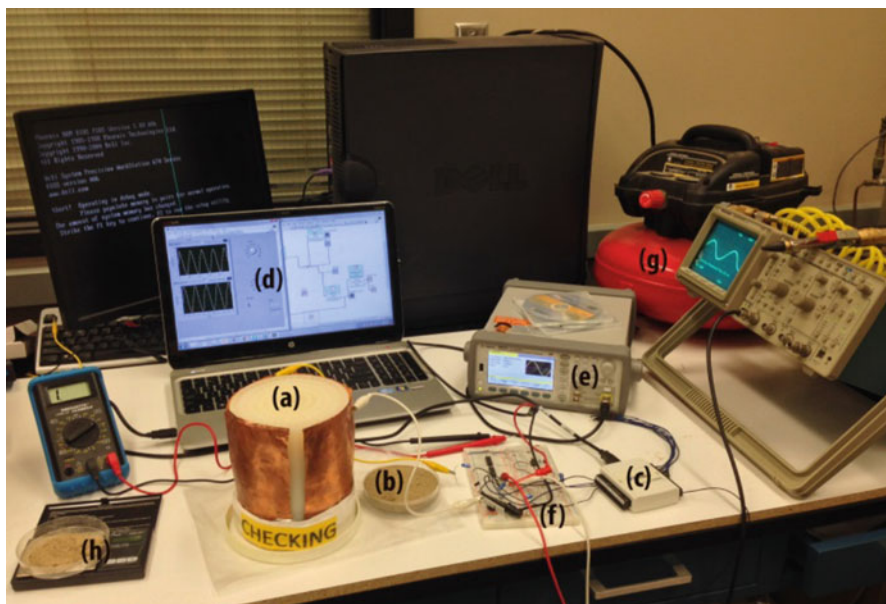| Number of electrodes	Independent measurement	Typical speed
6	15	400 [27]
8	28	200 [28]
12	66	100 [29]
16	120	50 [30]

Fig. 11.10 Experimental setup; (**a**) filter model, (**b**) sand as soot replacement, (**c**) NI DAQ, (**d**) Data acquisition program, (**e**) Signal generator, (**f**) Data acquisition circuit, (**g**) blower, (**h**) weighing of soot material

electrodes have to be placed around the DPF. The actual sensor was designed with four electrodes with the same DPF dimensions. This sensor was used in the Mechatronics Research Lab to build and experimental setup with function generator, oscilloscope, NI data acquisition system, etc. Figure 11.10 shows the complete experimental setup to identify the relationship between output voltage and amount of accumulated material.

A common practice in selecting the length of the ECT sensor is that the length should be larger than the diameter in order to avoid serious fringe effect [13]. So if the DPF diameter size is larger than the length, then fringe effect cannot be ignored. For experimental purpose in this chapter the filter model used with typical dimensions of 130 mm (5 in.) diameter and 152 mm (6 in.) length [13]. For the experimental setup length of the electrode is larger than the diameter of ECT system. For this experimental setup fringe effect completely ignored.

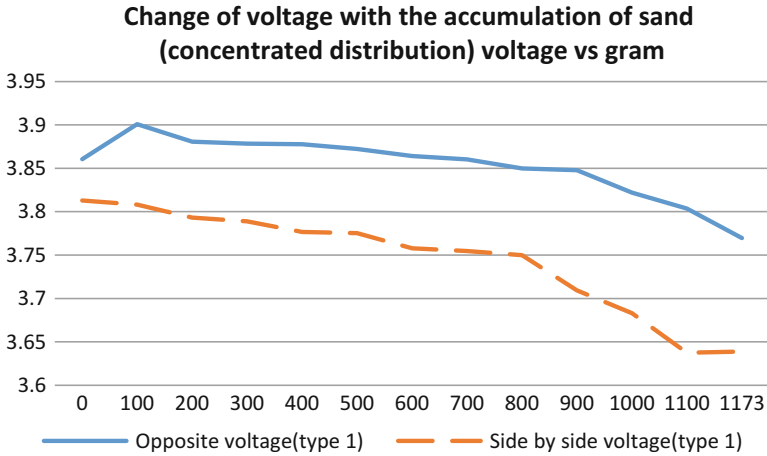**Change of voltage with the accumulation of sand (concentrated distribution) voltage vs gram**

Fig. 11.11 Change of voltage with the accumulation of sand (concentrated distribution)

The test bench is equipped with an automatic National instruments data acquisition system NI DAQ-6008 for capturing the capacitance values. We used dry sand as a replacement of soot to verify the approach of using ECT to soot detection. Properties of dry sand are given below [20]:

1. Density 1.60–1.70 g/cm^3
2. Electrical Resistivity 1000–100,000 Ω-cm
3. Specific Heat Capacity 0.753–0.799 J/g-°C
4. Thermal Conductivity 0.270–0.340 W/m-K

Two different methods of sand distribution have been considered while conducting the experiments. In concentrated distribution the assumption is at first sand start accumulation near to one particular electrode, and later filled up the whole filter. Voltage applied on the RC circuit is 4 V with 100 kHz frequency, approximation of frequency done from previous studies and experiments [21], which resulted output voltage plot as in Fig. 11.11. On the other hand in uniform distribution the assumption is the sand distributed equally through the whole filter resulted output voltage plot shown in Fig. 11.12. Data acquisition hardware senses the change of sand accumulation by the changes of the capacitance plate voltages. Voltage signals were processed in LABVIEW and Figs. 11.11 and 11.12 illustrate the fact that with the increase of sand accumulation the voltage is also changing.

From Figs. 11.11 and 11.12, it is evident that with the change of soot deposition output capacitor voltage also occurs in a detectable range. To clarify the repeatability of the results, another graph is shown in Fig. 11.13 which shows all the experimental results for change in voltage V_{AB} vs. soot deposition inside DPF from 10 to 100 %.
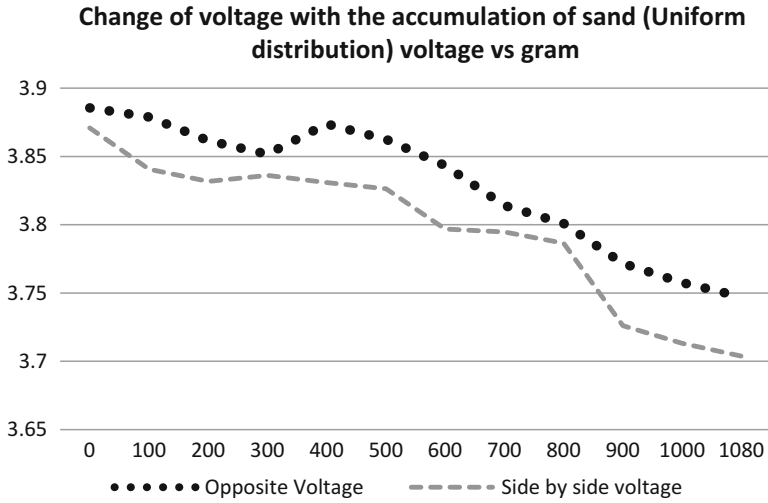
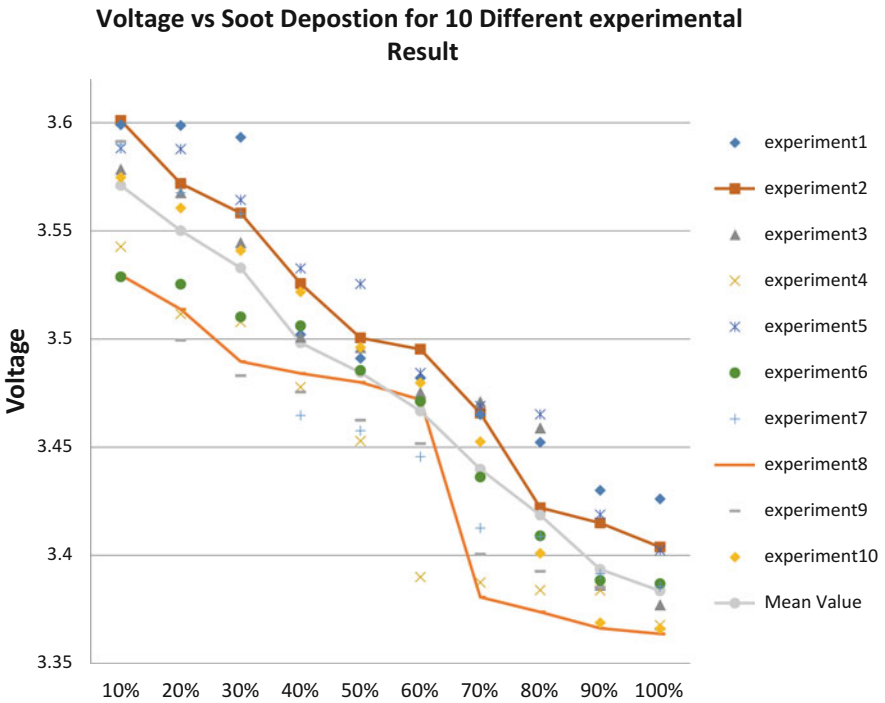**Fig. 11.12** Change of voltage with the accumulation of sand (Uniform distribution)

Fig. 11.13 Variability in experimental results

4 Experimental Results

4.1 Primary Deposition Detection Image Using Rule Based Method

A complete set of experiments has been conducted on the test bench using sand (density 1.6–1.9 g/cm^3) as soot model, Nylon 66 as DPF model, and four copper electrodes. Electrodes are arranged as shown in Fig. 11.14.

Sand has been poured into the grooves of DPF in a uniform fashion and simultaneously capacitance voltages between A-B V_{AB}, A-C V_{AC}, A-D V_{AD}, B-C V_{BC}, B-D V_{BD}, and C-D V_{CD} have been measured using the NI DAQ-6008 device and LABVIEW [22].

Now to prepare detection image we need to consider a 2 by 2 pixel matrix. To assign the value of each pixel we need to normalize the capacitance voltage value. Every pixel will have certain normalized capacitance voltage values as shown in Fig. 11.15.
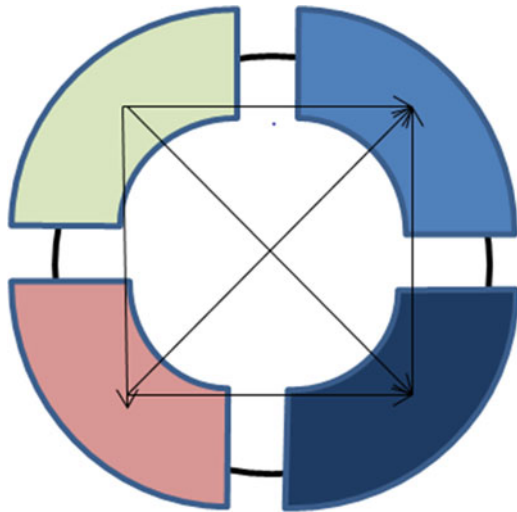
Fig. 11.14 Four electrode arrangement

Fig. 11.15 Assigned values of array

$V_{AB}+V_{AC}/4+V_{BD}/4$	$V_{AD}+V_{AC}/4+V_{BD}/4$
(A)	(B)
$V_{BC}+V_{AC}/4+V_{BD}/4$	$V_{CD}+V_{AC}/4+V_{BD}/4$
(C)	(D)

Table 11.2 Experimental capacitance data set

| Weight of sand	Total	A-B (V_{AB})	A-C (V_{AC})	A-D	B-C	B-D	C-D
0	0	3.946	3.998	3.902	3.857	3.963	3.823
100	100	3.983	3.961	3.897	3.806	3.902	3.851
100	200	3.972	3.957	3.889	3.904	3.894	3.802
100	300	3.966	3.955	3.878	3.838	3.876	3.828
100	400	3.925	3.942	3.878	3.826	3.867	3.822
100	500	3.942	3.941	3.867	3.821	3.856	3.823
100	600	3.929	3.91	3.856	3.866	3.87	3.813
100	700	3.905	3.904	3.823	3.848	3.858	3.733
100	800	3.867	3.928	3.849	3.838	3.878	3.722
100	900	3.743	3.929	3.849	3.88	3.916	3.747
100	1000	3.636	3.915	3.847	3.878	3.893	3.743
100	1100	3.503	3.879	3.832	3.841	3.865	3.696
100	1200	3.536	3.871	3.823	3.784	3.839	3.69
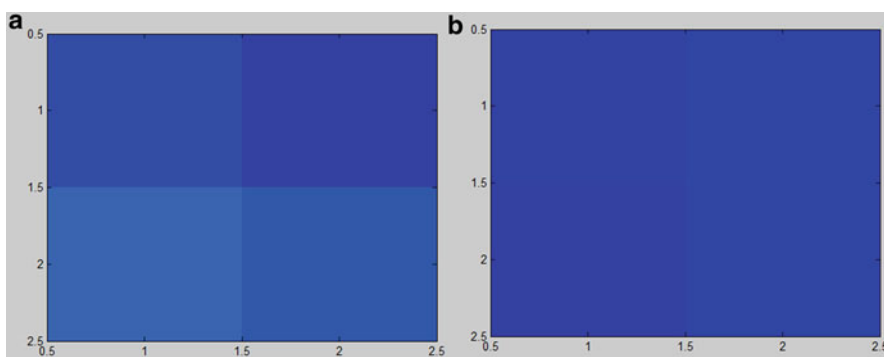
Fig. 11.16 Rule based tomographic image of DPF with zero and 25 % fill. (**a**) DPF zero fill; (**b**) DPF 25 % fill

Based on the mathematical relationship shown in Fig. 11.15, each pixel value can be evaluated. A scenario where DPF model is 25 % full means that the total fill is 300 g. According to Fig. 11.15 and voltage values from Table 11.2 pixel values of [A] pixel are calculated below.

[A] Pixel value = VAB + VAC/4 + VBD/4 = 3.9656 + (3.9549/4) + (3.87556/4) = 5.923215

Using this pixel values four detection images have been generated. Figures 11.16 and 11.17 show different stages of soot deposition images.

In both Figs. 11.16 and 11.17, darkness of the image correlates with noticeable changes in the material deposition.
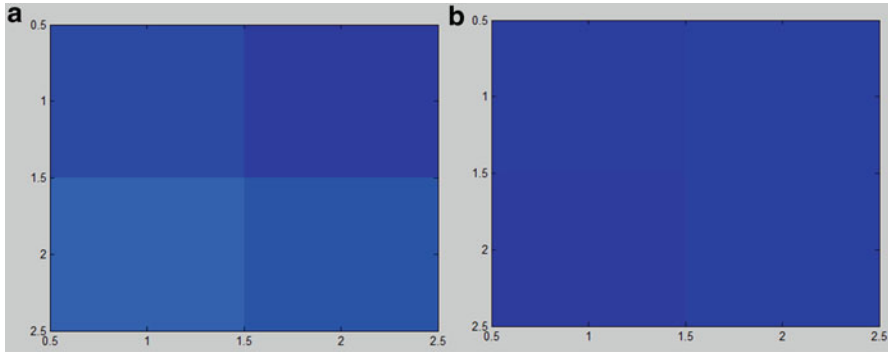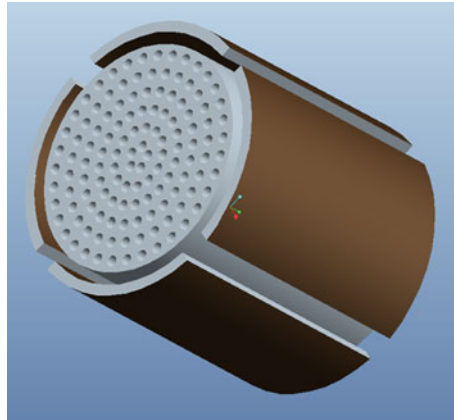
Fig. 11.17 Rule based tomographic image of DPF with 50 and 75 % fill. (**a**) DPF 50 % fill; (**b**) DPF 75 % fill

Fig. 11.18 DPF with four capacitor plates

4.2 Linear Back Projection Based Tomographic Image Generation

After generating primary soot detection image, the LBP method was used to generate tomographic image which is a complex algorithm compared to the previous method. To create a tomographic image using LBP method, sensitivity matrix was created [23]. The complete set of measured interelectrode capacitance values is required to reconstruct one permittivity distribution image. Figure 11.19 shows a 4×4 square pixel grid used to display the permittivity distribution image of a 4-electrode sensor, distributed as Fig. 11.18, having circular intersection of DPF. From this (4×4) square pixel grid containing 16 pixels, all 16 are needed to construct the cross-sectional image of the DPF.

A proper sensitivity map, purpose of which is to aid in selecting the proper pixel that individually contributes to the capacitance changes, has been developed in Table 11.3 for visualizing the electric field established between two electrodes when one of them is excited.
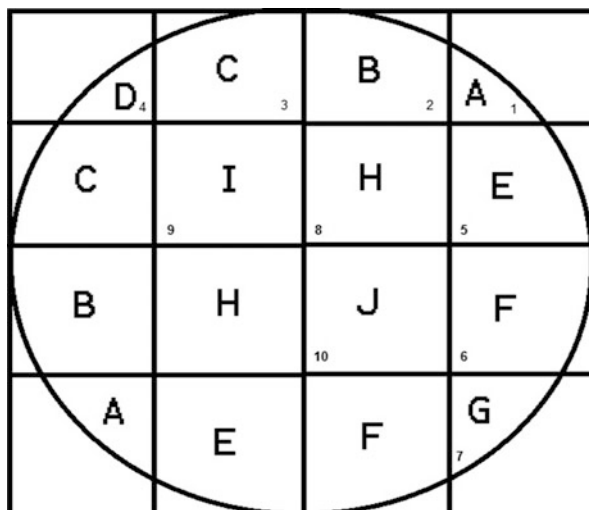
Fig. 11.19 Square pixel grid

Table 11.3 Sensitivity map

Pixel	A-B	A-C	A-D	B-C	B-D	C-D
A	3.692	3.638	3.550	3.670	3.641	3.944
B	3.685	3.635	3.556	3.662	3.680	3.830
C	3.554	3.657	3.600	3.644	3.649	3.868
D	3.537	3.578	3.669	3.637	3.671	3.487
E	3.540	3.644	3.593	3.582	3.657	3.884
F	3.555	3.678	3.641	3.664	3.717	3.934
G	3.597	3.599	3.539	3.562	3.615	3.893
H	3.528	3.645	3.642	3.643	3.687	3.971
I	3.659	3.623	3.577	3.651	3.678	3.976
J	3.509	3.638	3.636	3.671	3.710	3.975
A1	3.509	3.657	3.625	3.687	3.744	3.968
B1	3.702	3.666	3.618	3.646	3.719	3.591
C1	3.716	3.678	3.594	3.710	3.718	3.935
E1	3.535	3.646	3.564	3.688	3.698	3.993
F1	3.713	3.667	3.588	3.709	3.716	3.992
G1	3.626	3.657	3.569	3.658	3.661	3.970

Sensitivity map is one of the crucial parts of a tomographic image generation in LBP method. Sensitivity map needed to be created only for one calibration. For both end of the range of soot deposition to be measured the data in zero fill and then the data in complete fill have been obtained. These high and low data sets have been used to calibrate the sensor.

Once the sensor has been calibrated, capacitance data during various loading cycles has been obtained and normalized. The method of normalization took place in single step. As for example one set data normalized has been shown below.

Table 11.4 Normalized sensitivity map

	A-B	A-C	A-D	B-C	B-D	C-D
A	0.928425	0.715134	0.186846	0.570407	0.348201	0.896534
B	0.912299	0.693707	0.217141	0.542215	0.557531	0.682336
C	0.615689	0.843042	0.445987	0.4767	0.393433	0.753609
D	0.577745	0.308856	0.804782	0.449124	0.51081	0.040341
E	0.584872	0.757605	0.408925	0.250021	0.435328	0.784783
G	0.619509	0.988106	0.657259	0.548826	0.760222	0.878138
H	0.033826	0.454527	0.131434	0.177413	0.209671	0.801636
I	0.556456	0.760303	0.664233	0.47097	0.598349	0.947324
J	0.37693	0.09076	0.063197	0.137947	0.101703	0.510613
A1	0.514518	0.71374	0.63178	0.572083	0.722203	0.954597
B1	0.514986	0.844774	0.577853	0.632677	0.90279	0.942237
C1	0.952372	0.90748	0.542103	0.48249	0.770144	0.234994
E1	0.983385	0.986083	0.416882	0.712909	0.764726	0.879863
F1	0.572664	0.769117	0.261937	0.635198	0.654436	0.989346
G1	0.975881	0.908155	0.385068	0.712414	0.753763	0.987084

$$S_i(N) = \frac{C_i - C_i(\text{emp})}{C_i(\text{full}) - C_i(\text{emp})} \quad \text{for} \quad i = 1 \ldots M.$$

C_i (empty) $= 3.82\ V$ and C_i (full) $= 3.25\ V$

From Table 5.2 a particular $C_i = 3.55\ V$

So $S_i = (3.55 - 3.82) / (3.25 - 3.82)$

$S_i = 0.47 \quad (0 < S_i < 1)$

Based on this calculation Table 11.4 normalized sensitivity map has been generated using Table 11.3.

After assembling the sensitivity map, it was time to detect the change in material deposition inside DPF. How the soot has been distributed inside a DPF is still not been confirmed by any method. All diesel engine experts suggested that primarily soot deposited more densely near to the outer rim of the DPF because of the high speed entry into a DPF. Based on that approximation to measure soot deposition inside DPF, two different approaches have been selected:

(a) 10 % increment in soot load and
(b) 20 % increment in soot load.

Based on these approaches model DPF has been loaded with respect to time. After the completion of soot deposition inside the DPF, one set of capacitance values has been measured. This set of experimental values after normalization multiplied with previously found normalized sensitivity map which brings us the pixel values according to LBP. The pixel values in the permittivity images are similarly normalized using the high and low pixel values so they have 0 for the lower value and 10 for the higher value. The permittivity distribution of the soot loading along circumference inside a DPF with 10 % increment has been shown in Fig. 11.20.
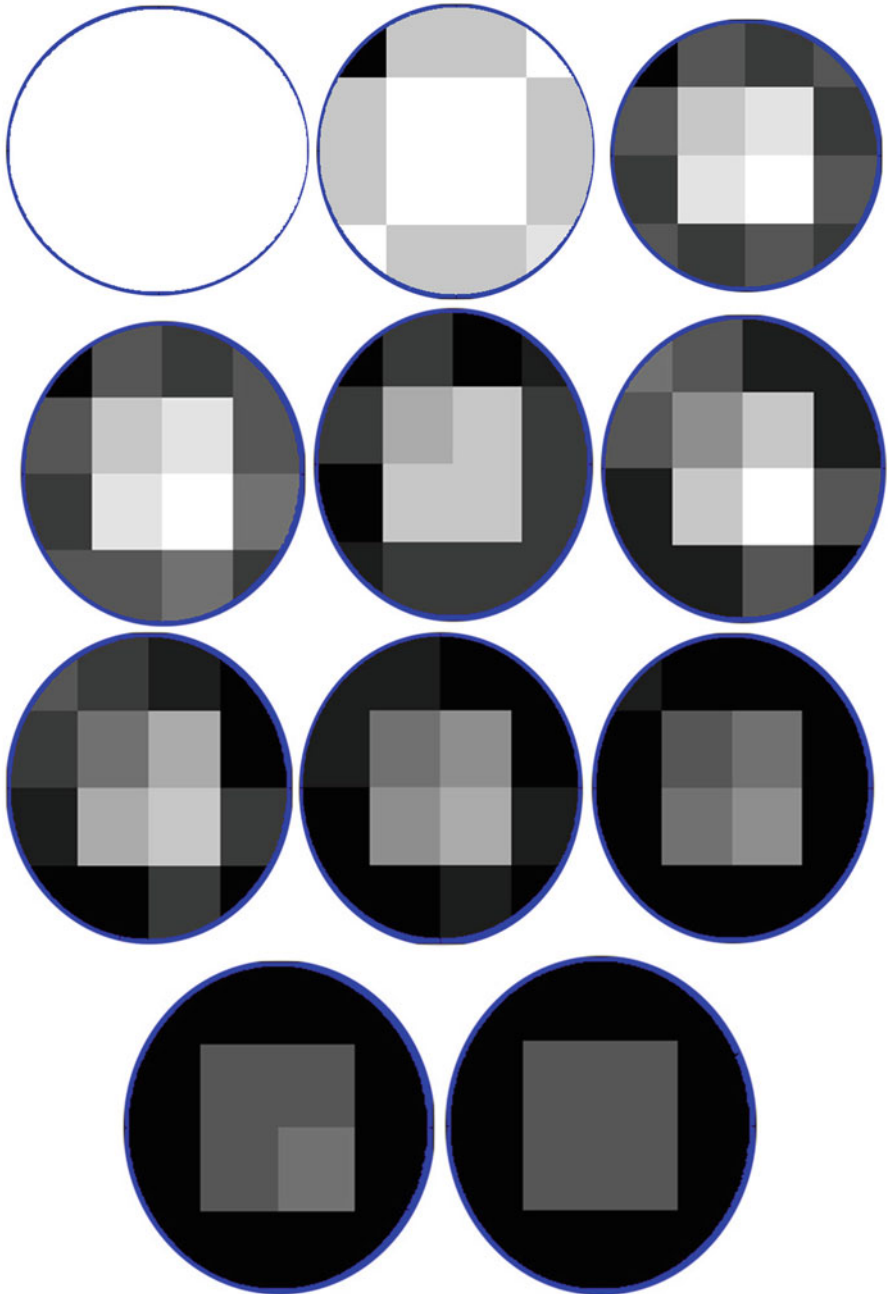
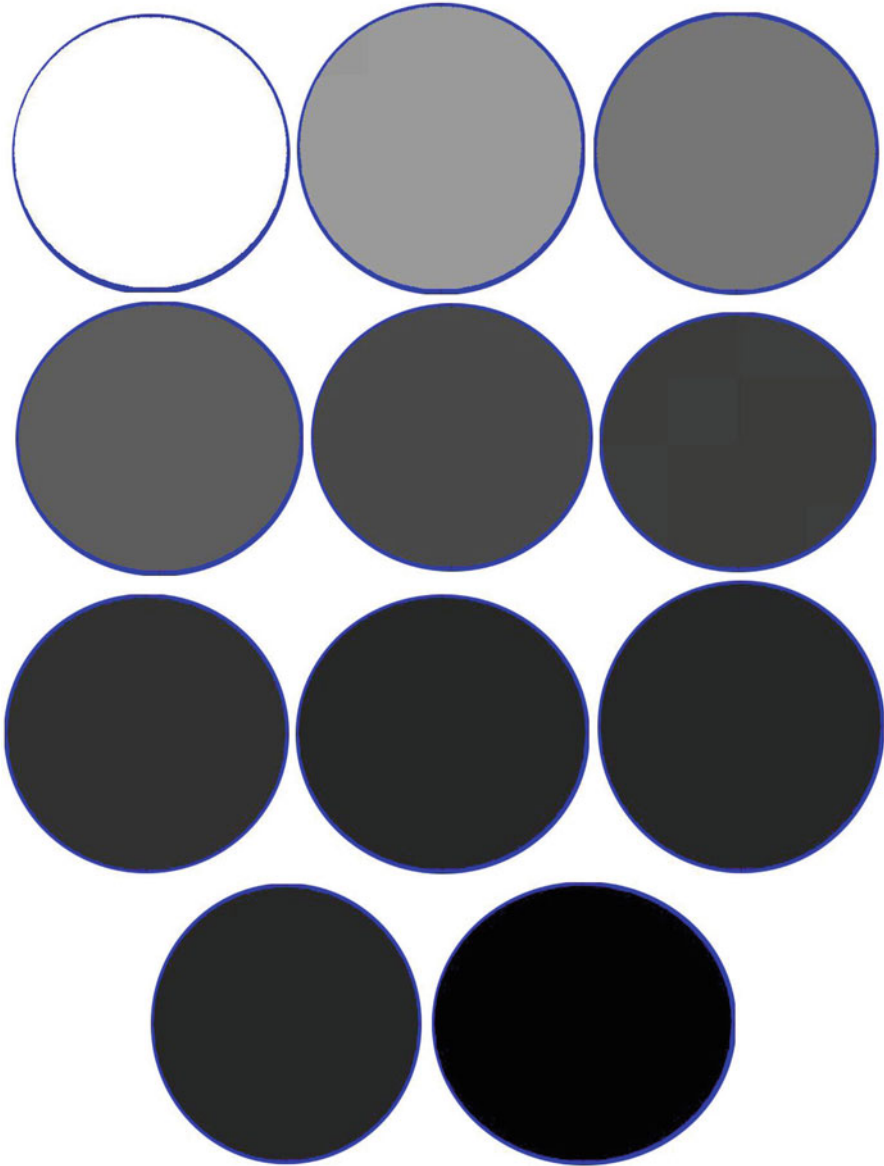**Fig. 11.20** Tomographic images for 10 % increment along circumference

Fig. 11.21 Tomographic images for 10 % increment full

The permittivity distribution of the full soot loading inside a DPF with 10 % increment has shown different tomographic images than the soot loading along circumference. New tomographic images have been shown in Fig. 11.21.
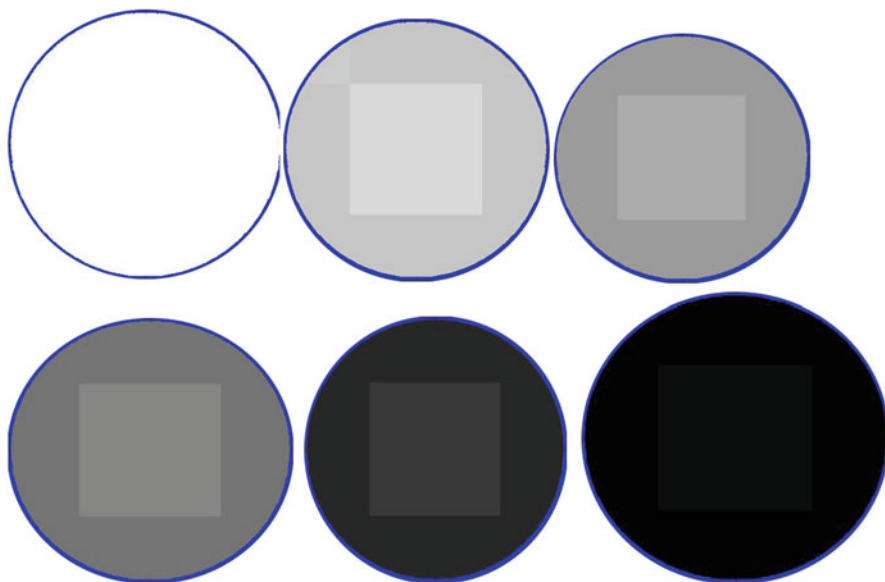
Fig. 11.22 Tomographic images for 20 % increment full along circumference loading

The permittivity distribution of the soot loading along circumference inside a DPF with 20 % increment has been shown in Fig. 11.22.

The permittivity distribution of the full soot loading inside a DPF with 20 % increment has been shown in Fig. 11.23.

From Figs. 11.20, 11.21, 11.22, and 11.23 DPF tomographic images for different loading condition have been recorded. It's evident that tomographic images are capable of capturing the soot deposition amount change. In Figs. 11.20 and 11.21, tomographic images of last few stages show similar color concentration, because at those stages the total mass of soot inside the DPF was so high that the detection voltage did not vary significantly, as can be seen from Figs. 11.11 and 11.12 plots.

4.3   Rule Based Tomographic Image Using PRINTEX U Material

Printex U carbon black material has always been used as a replacement of diesel soot for experimentation with diesel particulate emission in the industry. After generating primary tomographic images, printex u was used in the experimental setup to establish the effectiveness of the technology in case of soot. In Table 11.5 the output voltages in ECT electrodes have been shown using printex u.

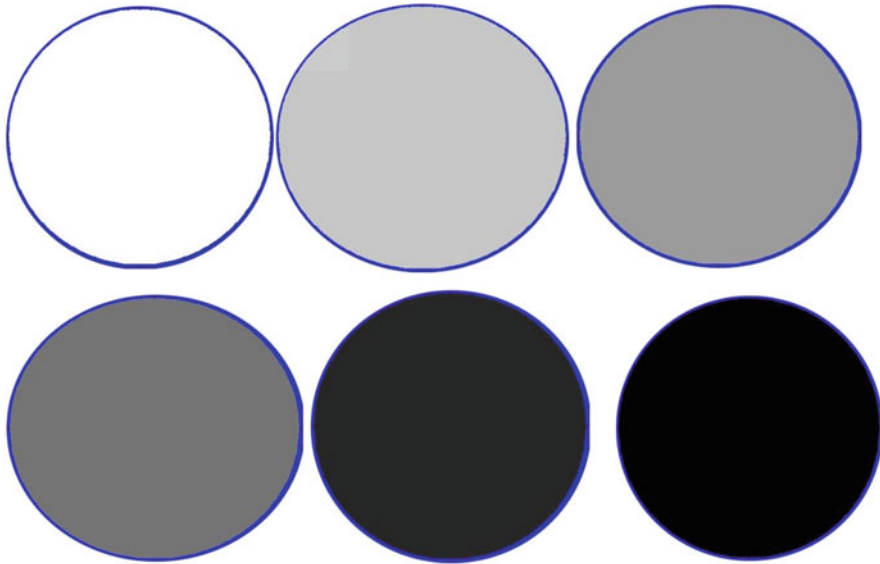From Table 11.5 tomographic images of DPF have been generated using the rule based method of Fig. 11.15.

Fig. 11.23 Tomographic images for 20 % increment full loading

Table 11.5 Output voltage using printex u

Weight (g)	A-B	A-C	A-D	B-C	B-D	C-D
0	3.6294	3.7088	3.6825	3.7547	3.7865	3.6789
18.7	3.5779	3.6219	3.5363	3.6439	3.6878	3.5933
18.7	3.3428	3.4265	3.3828	3.5889	3.6612	3.5668
18.9	3.2675	3.3525	3.3658	3.4256	3.5254	3.4864
19	3.2675	3.3665	3.1920	3.3738	3.3451	3.2780
18.8	3.1747	3.3279	3.1137	3.1090	2.9336	3.1940
19.4	2.7117	3.2547	3.0503	2.8826	2.8642	3.1044
19.9	2.2951	3.0660	2.9979	2.8242	2.8516	3.0916
20.3	2.2853	3.0369	2.9934	2.8196	2.8501	3.0782
19.4	2.2708	3.0202	2.9763	2.7988	2.8430	3.0775
18.9	2.2238	2.8842	2.8527	2.6458	2.7083	2.9343
20	2.2192	2.5984	2.4190	2.5135	2.5438	2.6997

From Fig. 11.24 tomographic images it is evident that presence of soot inside a DPF can be detected using rule based method but location of soot deposition is not conclusive, because instead of uniform soot deposition tomographic image shows high concentration of soot deposition in one particular quadrant which is quite unlikely. So it's evident that to detect soot location with soot detection a more robust method has to be implemented, at that point again LBP method comes in. LBP helps to create tomographic images of DPF with the help of sensitivity pixels so location of soot deposition is also possible.
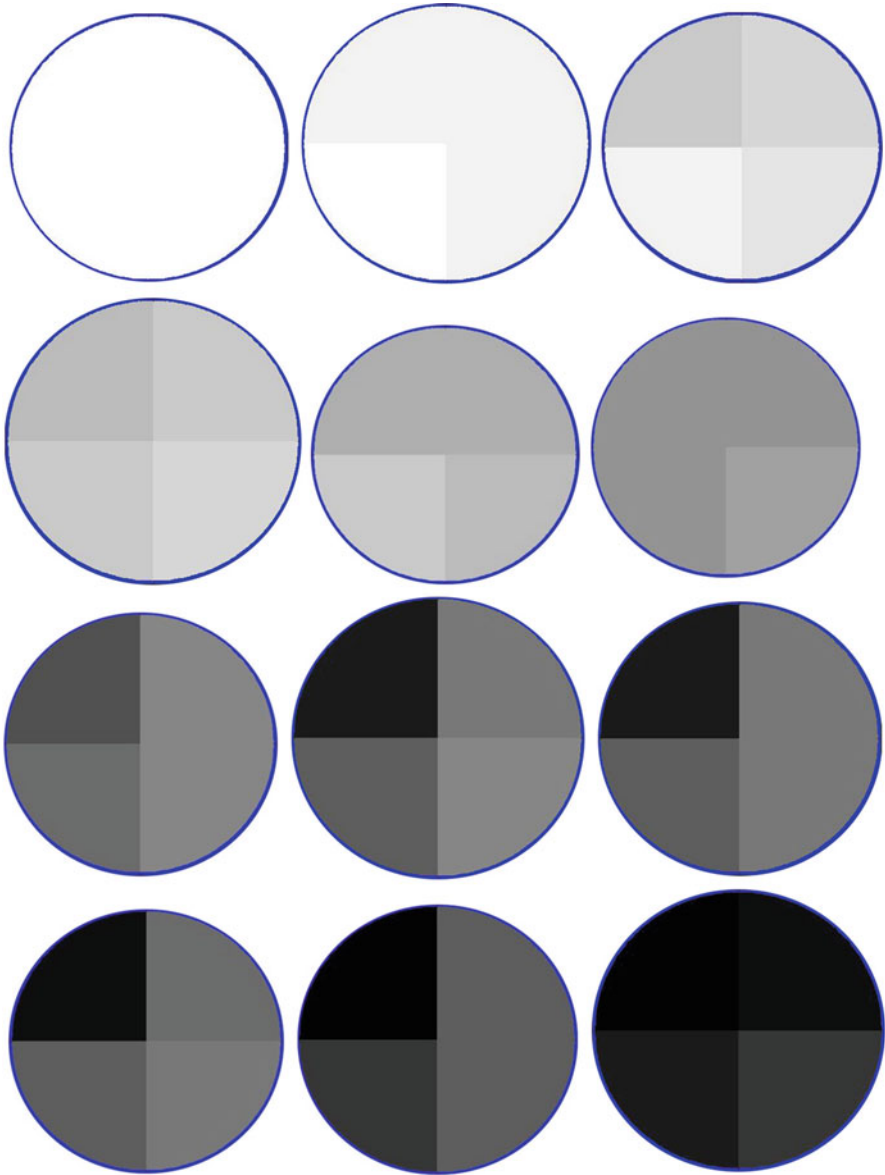
Fig. 11.24 Rule based tomographic image using printex u

4.4 Linear Back Projection Based Tomographic Image Using PRINTEX U

Though dry sand is not very close of original soot but for experimentation use of dry sand increases the flexibility. Eventually when Printex U has been used as a soot replacement, then more vivid voltage difference is visible compared to dry sand. In Fig. 11.25, all electrode voltage differences between soot and dry sand are shown.

From Fig. 11.25, all voltage output comparison it is visible that voltage changes due to Printex U deposition inside the DPF are more vivid than dry sand deposition. Usually dry sand addition causes voltage change range from 3.95 to 3.75 V. On the other hand Printex U deposition causes voltage change range to shift to 3.7–2.2 V, though with the increase of the range linearity of the output voltage diminishes. This increase of voltage range causes better data acquisition and more versatile range of soot deposition can be recorded.

Comparing Figs. 11.20, 11.21, 11.22, and 11.23 (dry sand tomographic image) with Figs. 11.26 and 11.27 (printex U tomographic image), it is evident that due to bigger voltage output range Printex u deposition can give much better image contrast during tomographic image generation. Again comparing Figs. 11.26 and 11.27 with Fig. 11.24 (rule based image formation), contrast between pixels shows that LBP method is more accurate method.

4.5 Statistical Error Calculation in Output Voltage

In order to verify the accuracy in the result or to check the repeatability, statistical analysis has been conducted on the experimental results. For dry sand and 10–90 % soot load conditions, all the mean, standard deviation, and standard error for experimental voltage V_{AB} are shown in Table 11.6.

Table 11.6 points to the fact that experimental results show high repeatability with a very low standard error. Following same steps, the mean, standard deviation, and error for experimental results using Printex U material (output voltage V_{ab}) are shown in Table 11.7.

The repeatability of the recorded data was also checked by observing the frequency distribution for a certain result range. The frequency distribution for voltage output data for dry sand and printex u materials has been plotted in Figs. 11.28 and 11.29.

Figures 11.28 and 11.29 show very small irregularities in the result of $V_{ab.}$ Frequency of the voltage values varies, and that variation causes the irregularities in voltage vs deposition curve like Figs. 11.10 and 11.11 or in the tomographic images.
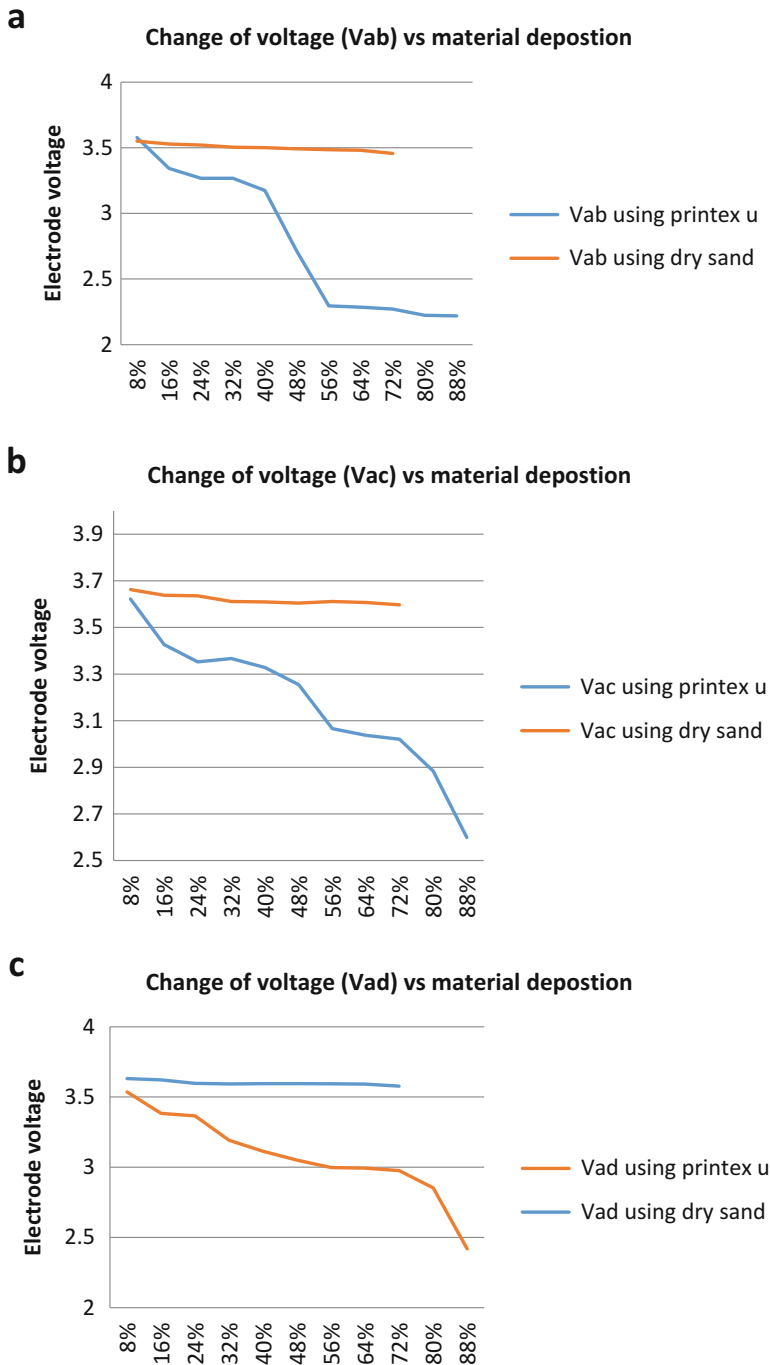
a

b

c

Fig. 11.25 (**a–f**) Represents all set of n of electrode voltage differences between Printex u and dry sand

d

Change of voltage (Vbc) vs material depostion

e

Change of voltage (Vbd) vs material depostion

f

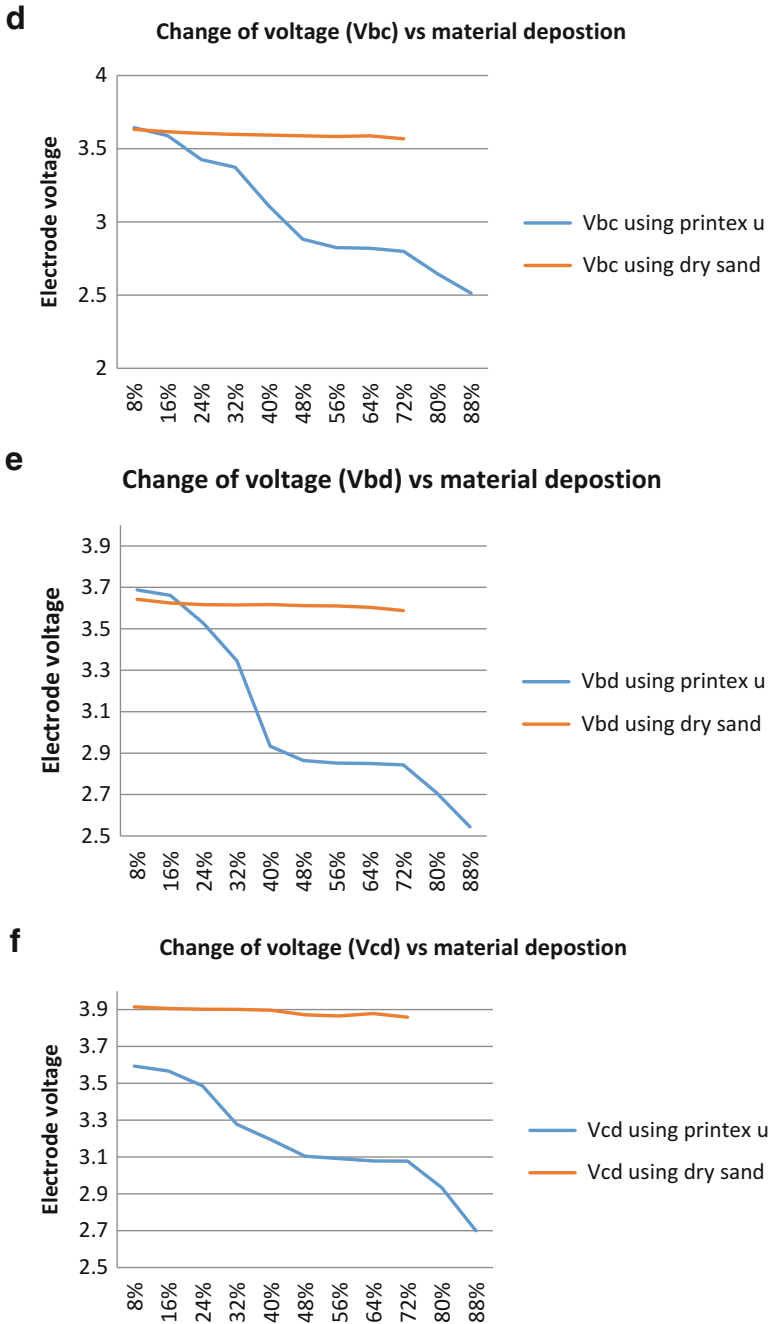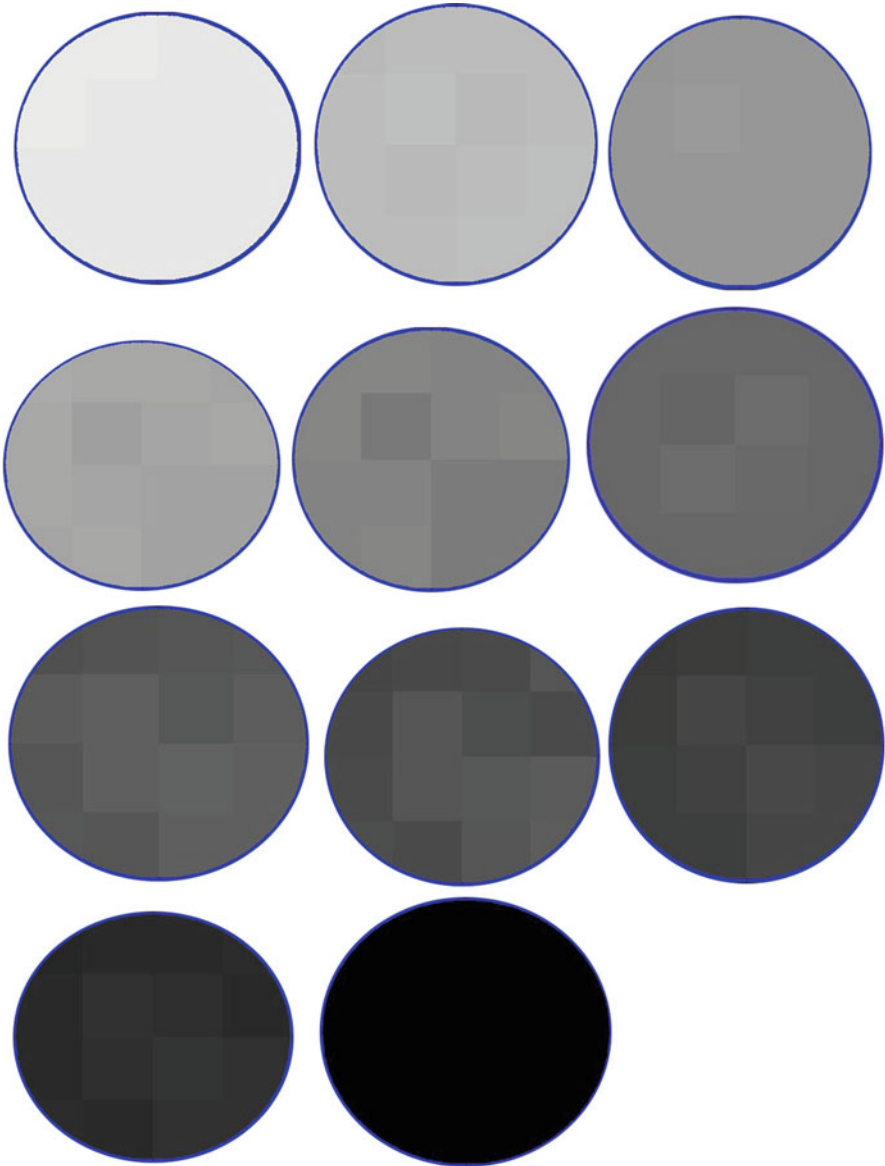**Change of voltage (Vcd) vs material depostion**

Fig. 11.25 (continued)

Fig. 11.26 Tomographic image using printex U full loading
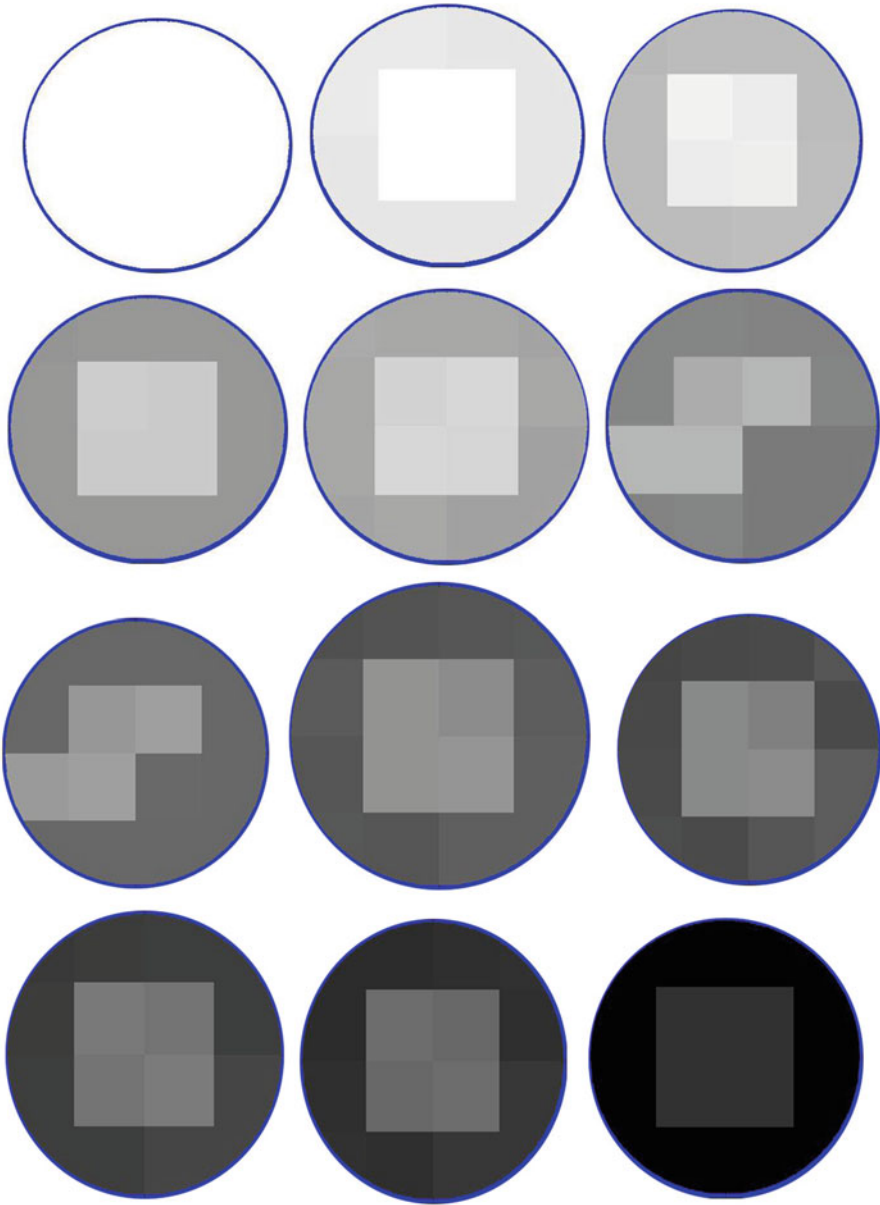
Fig. 11.27 Tomographic image using printex U circumferential loading

Table 11.6 Standard deviation and error

Material deposition %	Mean	SD	Standard error
10	3.572	0.028	0.009
20	3.550	0.035	0.012
30	3.535	0.036	0.012
40	3.499	0.023	0.008
50	3.485	0.022	0.007
60	3.465	0.030	0.010
70	3.434	0.036	0.012
80	3.417	0.032	0.011
90	3.393	0.021	0.007

Table 11.7 Standard deviation and error

Material deposition %	Mean	SD	Standard error
9	3.514	0.057	0.018
18	3.265	0.204	0.065
27	3.182	0.213	0.067
36	3.075	0.206	0.065
45	2.952	0.200	0.063
54	2.854	0.237	0.075
63	2.751	0.215	0.068
72	2.642	0.170	0.054
81	2.493	0.150	0.047
90	2.356	0.136	0.043
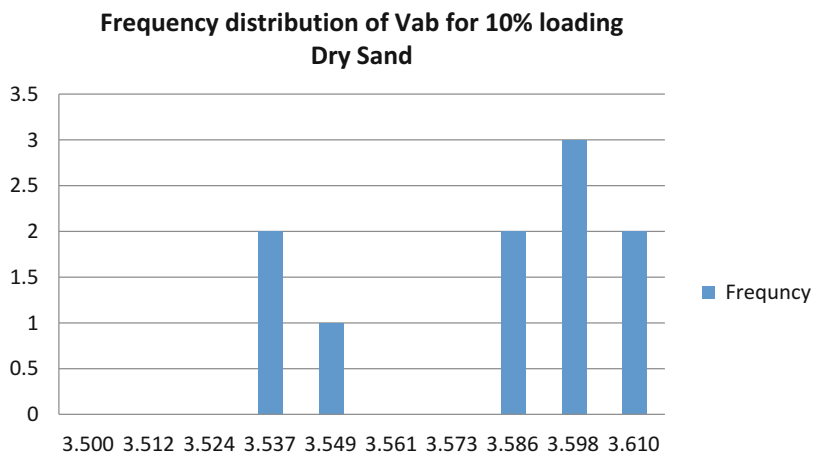
Fig. 11.28 Frequency distribution for dry sand

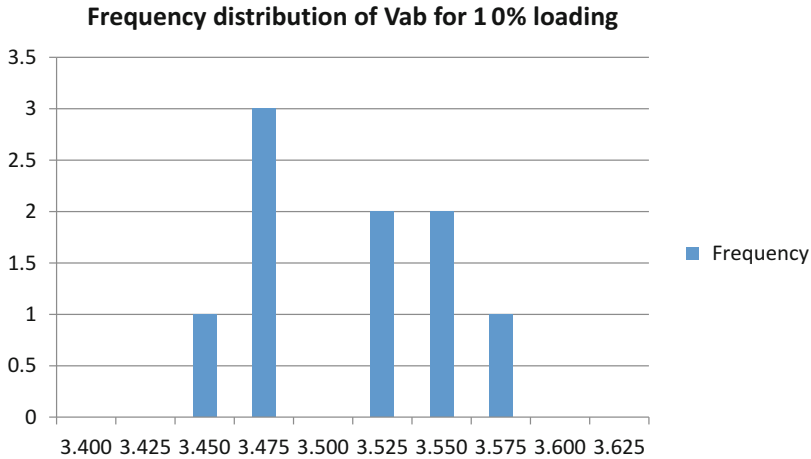**Frequency distribution of Vab for 10% loading**

Fig. 11.29 Frequency distribution for Printex U

5   Conclusion

DPF is a very important inclusion in a diesel engine to maintain the particulate emission level below federal regulatory limits. Though there are several types of DPF technologies available in the market, this work focuses on the model closest to cordierite diesel filter. Most of the DPF currently available in the market can reduce PM emission by almost 99 %. After a number of cycles, the PM or soot accumulates in the DPF which must be removed from DPF to prevent high backpressure or event microscopic damage in the filter substrate. An active regeneration process burns the soot by adding fuel in the exhaust which generates very high temperatures at DPF. Such high temperatures can cause cracks inside the DPF surfaces and excessive fuel use can cause fuel penalty. To maintain an optimized amount of fuel for soot burning, accurate knowledge of deposited soot is necessary. To maintain both DPF structural integrity and emission regulation, it is very important to know the exact amount soot deposited inside a DPF. There are several technologies available to detect and measure the soot load status inside a DPF, e.g., pressure drop sensor, particulate emission sensor, microwave sensor, and radio frequency sensor. In this work, we presented a novel soot sensing sensor based on ECT. The patented design [24] utilizes ECT methodology for DPF soot measurement where a number of electrodes surround the DPF outer shell. Any soot or PM deposition inside the DPF will cause a change in dielectric properties inside the DPF. This change of dielectric property will cause a change in the capacitance. Electrodes of ECT system act a set of capacitors surrounding the DPF. These electrodes capture a combination of capacitance data and voltage data. Two different methodologies have been investigated to generate the tomographic images: (a) rule based method and (b) LBP method.

In this work, soot replacement Printex U has been used because it has the closest dielectric property of original soot. The ECT system for soot load measurement in a model DPF was built with Nylon 66 as substrate material and four copper electrodes. Initially dry sand was used as soot material due to its easy availability and cleaner handling properties. With encouraging results with sand as a soot material, the Printex U material (having closest chemical properties of soot originally generated from an engine) has been used in the experimentations. Data obtained from Printex U experiments showed improved sensitivity of the voltage output with soot load when compared to dry sand results. The whole range of experiments were repeated using Printex U material which resulted in vibrant tomographic images with more contrast.

The experimental results show that ECT based soot load sensing system is responsive to the change of soot material accumulation in DPF. The final tomographic images of both dry sand and Printex u materials showed that detection of soot deposition inside a DPF is a viable method for soot load measurement using ECT based methodology.

References

1. Twigg MV, Phillips PR (2009) Cleaning the air we breathe—controlling diesel particulate emissions from passenger cars. Platin Met Rev 53(1):27–34
2. Adler J (2005) Ceramic diesel particulate filters. Int J Appl Ceram Technol 2(6):429–439
3. Fischerauer G, Forster M, Moos R (2009) Sensing the soot load in automotive diesel particulate filters by microwave methods. Meas Sci Technol 21:235–247
4. Sappok A, Bromberg L, Parks J, Prikhodko V (2010) Loading and regeneration analysis of a diesel particulate filter with a radio frequency-based sensor. SAE technical paper, San Diego, CA
5. Ohyama N, Nakanishi T, Daido S (2008) New concept catalyzed DPF for estimating soot loadings from pressure drop. SAE technical paper 2008-01-0620
6. Singh N, Rutland C, Foster D, Narayanaswamy K et al (2009) Investigation into different DPF regeneration strategies based on fuel economy using integrated system simulation. SAE technical paper 2009-01-1275
7. Rose D, Boger T (2009) Different approaches to soot estimation as key requirement for DPF applications. SAE technical paper, Detroit, MI. doi:10.4271/2009-01-1262
8. Husted H, Roth G, Nelson S, Hocken L, Fulks G, Racine D (2012) Sensing of particulate matter for on-board diagnosis of particulate filters. SAE Int J Engines 5(2):235–247
9. Strzelec A, Bilheux H, Finney C, Daw C et al (2009) Neutron imaging of diesel particulate filters. SAE technical paper, San Antonio, TX
10. Boylestad RL (2007) Capacitor. In: Introductory circuit analysis, 11th edn. Pearson Prentice Hall, New Jersey. Chapter 10, Section 10.3, pp 398–410
11. Donthi SS (2004) Capacitance based tomography for industrial applications. M.Tech. Thesis, Electrical Engineering Dept., IIT, Bombay
12. Koledintseva MY, DuBroff RE, Schwartz RW (2006) A Maxwell Garnett model for dielectric mixtures containing conducting particles at optical frequencies. Prog Electromagn Res 63: 223–242
13. Cao Z, Xu L, Fan W, Wang H (2011) Electrical capacitance tomography for sensors of square cross sections using Calderon's method. IEEE Trans Instrum Meas 60(3):900–907

14. Wuqiang Y (2010) Design of electrical capacitance tomography sensors. Meas Sci Technol 21(4):42001–42013
15. Gamio JC, Ortiz-Alemán C, Martin R (2005) Electrical capacitance tomography two-phase oil-gas pipe flow imaging by the linear back-projection algorithm. Geophys Int 44(3):265–273
16. Process Tomography Ltd (2009) Electrical capacitance tomography system Type TFLR5000 operating manual, vol 1. Fundamentals of ECT. Process Tomography Ltd., England
17. Matweb. Material property data. http://www.matweb.com/
18. Ferro Ceramic Grinding, Inc. http://www.ferroceramic.com/
19. Pryuor L, Sclobohm R, Brownell B (2009) A comparison of aluminum vs. copper as used in electrical equipment. GE Consumer & Industrial website https://www.geindustrial.com/sites/geis/files/gallery/
20. Berney ES, Smith DM (2008) Mechanical and physical properties of ASTM C33 sand [Type of medium]. http://www.dtic.mil/. Feb 2008
21. Domenick B (1955) A graphical sinusoidal analysis of a nonlinear RC phase-shift feedback circuit. Proc IRE 43(6):679–684
22. Fountain T (1993) Software advances in measurement and instrumentation. In: IEE colloquium on software instrumentation—software components, Feb 1993, pp 1–45
23. Rerkratn A, Chitsakul K, Soisup A, Wuti V (2010) Electrical capacitance tomography system for monitoring process flow in pipe. In: Proceedings of SICE annual conference 2010, Taipei, China, pp 3229–3232
24. Huq R, Anwar S (2015) Real-time soot measurement in a diesel particulate filter. US Patent No. 9,151,205. 6 Oct 2015
25. Michel RP, Baican R, Schubert E (1993) Soot particle properties in the microwave range. In: Microwave conference, 1993. 23rd European, Madrid, Spain, pp 959–960
26. Pyzik AJ, Li CG (2005) New design of a ceramic filter for diesel emission control application. Int J Appl Ceram Technol 2(6):440–451
27. Waterfall RC, He R, White NB, Beck CM (1996) Combustion imaging from electrical impedance measurements. Meas Sci Technol 7:369–374
28. Yang WQ, Chondronasios A, Nguyen VT, Nattras S, Betting M, Ismail I (2004) Adaptive calibration of a capacitance tomography system for imaging water droplet distribution. Flow Meas Instrum 15:249–258
29. Yang WQ, Stott AL, Beck MS, Xie CG (1995) Development of capacitance tomographic imaging systems for oil pipeline measurements. Rev Sci Instrum 66:4326–4332
30. Yakowski T, Miko M, Vlaev D, Mann R, Follows GW, Boxman A, Wilson MPW (1999) Imaging nylon polymerisation processes by applying electrical tomography. In: Proceedings of the 1st World Congress on industrial process tomography, Buxton, UK, 14–17 Apr, pp 383–387

Chapter 12
Microfluidic Platforms for Bio-applications

Anas Alazzam, Bobby Mathew, and Saud Khashan

Abstract This chapter provides a brief overview of three actuation mechanisms that are relevant for biomedical applications of microfluidics. Actuation mechanisms are employed in the field of microfluidics for realizing unit operations such as focusing, switching, and separation. The topics dealt with in this chapter include dielectrophoresis, acoustophoresis, and magnetophoresis. The first section provides an introduction to these and related topics while the second section deals specifically on dielectrophoresis. The third and fourth sections detail acoustophoresis and magnetophoresis, respectively. This chapter concludes by providing a quick comparison of these different actuation methods.

Keywords Acoustophoresis • Dielectrophoresis • Magnetophoresis • Microfluidics • Micro total analysis systems • Point-of-care

1 Introduction

The field of microfluidics originated probably in the late 1970s when the first microfluidic device, a gas chromatography microdevice in silicon, was developed [1]. Since then the field of microfluidics has experienced an explosion of interest. This interest has led to the investigation of the fundaments of transport phenomena associated with microfluidic devices as well as application oriented development of the same. The level of interest in this field can be gauged from the presence of several journals solely dedicated to the same as well as the degree of funding from governmental as well as nongovernmental organizations. Microfluidic devices have found applications in several fields of engineering, namely biomedical, chemical, and mechanical engineering [2]. With regard to biomedical engineering,

---

A. Alazzam (✉) • B. Mathew
Department of Mechanical Engineering, Khalifa University, Abu Dhabi, UAE
e-mail: anas.alazzam@kustar.ac.ae

S. Khashan
Department of Mechanical Engineering, Jordan University of Science and Technology, Irbid, Jordan

applications include DNA amplification, label-free cell and biomolecule identification and capture/separation, immunoassay, cell culture, proteomics, plasmapheresis, and drug discovery [3, 4].

The quintessential characteristic of microfluidic devices is that they employ internal flow passages with hydraulic diameter smaller than 1 mm. The primary advantages of employing microfluidic devices include high surface area to volume ratio and enhanced heat and mass transfer coefficients [5, 6]. Other relevant merits of microfluidic devices include small footprint, low volume sample and reagent requirement, and portability [5–7]. All microfluidic devices operate within the laminar flow regime, i.e., Reynolds number lower than 2000; Reynolds number is mathematically stated in Eq. (12.1) [8]. With regard to microfluidic devices employed for biomedical applications, the Reynolds number is almost always smaller than unity [9].

$$Re = \frac{\rho D_{\text{hy}} \overline{V}_{\text{m}}}{\mu} \tag{12.1}$$

where Re (−) is the Reynolds number, ρ (kg m^{-3}) is the density of the medium, D_{hy} (m) is the hydraulic diameter of the microchannel, $\overline{V}_{\text{m}}$ (m s^{-1}) is the average velocity of the medium, and μ (Pa s) is the dynamic viscosity of the medium.

When Reynolds number is smaller than unity, the flow is specifically termed as Stokes flow [8]. In Stokes flow, the flow is fully developed which allows for neglecting the convective terms associated with Navier–Stokes equation in comparison with the viscous terms of the same and this in turn makes possible the development of an analytical equation for flow velocity (V_{m}) [8, 9]. The flow velocity and hydraulic diameter, for rectangular microchannels, are provided in Eqs. (12.2) and (12.3), respectively [10].

$$V_{\text{m}}(y,z) = 48 \frac{Q_{\text{m}}}{\pi^3 W_{\text{ch}} H_{\text{ch}}} \frac{\sum_{i=1,3,5}^{\infty} \left(\frac{(-1)^{\left(\frac{i-1}{2}\right)}}{i^3} \right) \cos\left[i\frac{\pi}{W_{\text{ch}}} \left(y - \frac{W_{\text{ch}}}{2} \right) \right] \left\{ 1 - \frac{\cosh\left[i\frac{\pi}{W_{\text{ch}}} \left(z - \frac{H_{\text{ch}}}{2} \right) \right]}{\cosh\left(\frac{i\pi}{2} \frac{H_{\text{ch}}}{W_{\text{ch}}} \right)} \right\}}{\left[1 - \frac{192}{\pi^5} \frac{W_{\text{ch}}}{H_{\text{ch}}} \sum_{i=1,3,5}^{\infty} \frac{\tanh\left(i\frac{\pi}{2} \frac{H_{\text{ch}}}{W_{\text{ch}}} \right)}{i^5} \right]} \tag{12.2}$$

$$D_{\text{hy}} = 2 \frac{W_{\text{ch}} H_{\text{ch}}}{(W_{\text{ch}} + H_{\text{ch}})} \tag{12.3}$$

where Q_m (m^3 s^{-1}) is the volumetric flow rate, H_{ch} (m) is the height of the microchannel, W_{ch} (m) is the width of the microchannel, y (m) is the distance of the location of interest along the width as measured from the bottom left corner, and z (m) is the distance of the location of interest along the height as measured from the bottom left corner. Equation (12.2) provides the velocity at a specific location over the cross section of the microchannel; since the flow is fully developed, there is no variation of velocity along the axial direction, i.e., x-direction.

In many instances, the width of the microchannel is much larger than the depth/height in which case velocity does not vary along the width as well, i.e., y-direction, except at the edges [8]. Subsequently, the velocity of flow in the microchannel can be written as shown in Eq. (12.4) [8]. Equation (12.4) provides the velocity at any height irrespective of the axial location.

$$V_m(z) = 6\frac{Q_m}{\alpha H_{ch}^2}\left(\frac{z}{H_{ch}}\right)\left(1 - \frac{z}{H_{ch}}\right) \quad \text{where} \quad \alpha = \frac{H_{ch}}{W_{ch}} \gg 1 \qquad (12.4)$$

where z (m) represents the distance of the location of interest along the height as measured from the bottom of the microchannel.

Microfluidic devices are primarily fabricated using techniques commonly encountered in the semiconductor industry [11]. Commonly employed microfabrication techniques include lithography, thin film deposition, wet and dry etching, and bonding [3, 4, 12]. With regard to magnetophoresis, electroplating is also an important microfabrication step [3, 4]. Several different substrates are used for realizing microfluidic devices [3, 4]. These include silicon, glass, SU8, poly methyl methacrylate (PMMA), and polydimethylsiloxane (PDMS) [3, 4]. The selection of the substrate material depends on the application. Microfabrication is capable of realizing microchannels with several different profiles; however, microfluidic devices for biomedical applications are almost always realized with square/rectangular profile.

Earlier it was mentioned that microfluidic devices are employed for label-free cell and biomolecule identification and capture/separation and immunoassays. For realizing these applications, it is necessary to subject cells and biomolecules to either attractive or repulsive force(s). This particular chapter is dedicated to provide an overview of three different actuation mechanisms widely employed in microfluidic devices, specifically dielectrophoretic, acoustophoretic, and magnetophoretic forces. Dielectrophoretic force is a type of electrical force which is based on the size and electrical permittivity and conductivity of the species while acoustophoretic force is based on the size, density, and compressibility of the species. On the other hand, magnetophoretic force is a magnetic force which depends on the size and magnetic susceptibility of the species. The second section of this chapter is dedicated to dielectrophoretic force while the third and fourth sections of this chapter are dedicated to acoustophoretic and magnetophoretic forces, respectively. Each section details the associated force as well as the approaches taken in implementing it for the abovementioned applications. With regard to microfluidic devices employed for capture/separation, the performance metrics include separation purity

and efficiency. Separation purity refers to the ratio of the number of desired species among all the species capture/separated while separation efficiency relates the number of desired species before and after executing capture/separate; these parameters are mathematically stated as in Eqs. (12.5) and (12.6).

$$\text{Purity} = \frac{\text{\# of targeted microscale entities in the collected sample}}{\text{\# of microscale entities in the collected sample}} \quad (12.5)$$

$$\text{Efficiency} = \frac{\text{\# of targeted microscale entities in the collected sample}}{\text{\# of targeted microscale entities in the original sample}} \quad (12.6)$$

2 Dielectrophoresis

Dielectrophoresis (DEP) is defined as the motion of polarizable microscale entities under the effect of nonuniform electrical field [13]. Microscale entities suspended in a conductive medium under similar conditions will experience a dielectrophoretic force whose magnitude and direction depend on the applied electric field and the properties of the microscale entities and their submersion medium [13]. DEP force exists due to the interaction between the nonuniform electric field and the induced dipole on the microscale entities located within the field.

DEP has been extensively used for biomedical applications [14, 15]. DEP-based microfluidic platforms have been demonstrated for cell separation, focusing, lysing, and transportation of cells and sub-cells [16–19]. DEP is one of the most promising techniques for point-of-care systems (POC) as it is label free, rapid, electrically manipulatable, noninvasive, and highly selective and sensitive. The recent progress in using DEP for cell separation holds promise for biomedical applications of microfluidic platforms [20, 21]. Several studies have showed that targeted cells, including yeast cells, lymphocytes, STEM cells, and cancer cells, can be successfully separated from a mixture of cells using DEP microdevices [14]. Moreover, there are strong evidences that DEP has little effect on cells [14, 22, 23]. The expression of DEP force acting on a spherical microscale entity is given by the following equation [16, 24]:

$$\boldsymbol{F}_{\text{DEP}} = \pi \varepsilon_0 \varepsilon_\text{m} R_\text{o}^3 Re\left[f_{\text{CM}}\right] \nabla \left(\boldsymbol{E} \cdot \boldsymbol{E}\right) \quad (12.7)$$

where the constants ε_o (F/m) and ε_m (−) are the vacuum permittivity and the suspending medium dielectric constant, respectively, R_o (m) is the radius of the microscale entity, $\boldsymbol{E}$ (V m^{-1}) represents the peak electric field, and $Re[f_{\text{CM}}]$ (−) is the real part of Clausius–Mossotti factor which is defined as [14]:

$$f_{\text{CM}} = \frac{\varepsilon_\text{p}^*(\omega) - \varepsilon_\text{m}^*(\omega)}{\varepsilon_\text{p}^*(\omega) + 2\varepsilon_\text{m}^*(\omega)} \quad (12.8)$$

where $\varepsilon_p^*(\omega)$ $(F\ m^{-1})$ and $\varepsilon_m^*(\omega)$ $(F\ m^{-1})$ are the frequency dependent complex permittivity of the particle and suspending medium, respectively. The complex permittivity, $\varepsilon^*(\omega)$, is a function of permittivity, ε, the angular frequency, ω (rad s^{-1}), and the electrical conductivity, σ (S m^{-1}), and is defined as:

$$\varepsilon^* = \varepsilon - j\frac{\sigma}{\omega} \tag{12.9}$$

The Clausius–Mossotti factor, f_{CM}, defined in Eq. (12.8) and the DEP force in Eq. (12.7) are valid only for spherical microscale entities with homogeneous properties. However, not all microscale entities are perfectly spherical or with uniform properties. Cells, viruses, and DNA are few examples of nonspherical with nonhomogeneous properties of microscale entities. Biological entities can be approximated as ellipsoid, cylinder, or any other shape. The DEP force and Clausius–Mossotti, f_{CM}, in this case need to be derived taking into consideration the shape and nonhomogeneity of properties.

Equations (12.8) and (12.9) indicate that the Clausius–Mossotti factor at high frequencies depends mainly on the dielectric constants of the microscale entity and medium while at low frequencies it is determined by the electrical conductivities of the microscale entity and medium. Equation (12.8) at high ($\omega \to \infty$) and low ($\omega \to 0$) angular frequencies can be written as:

$$f_{CM_{\omega \to \infty}} \cong \frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m} \tag{12.10}$$

$$f_{CM_{\omega \to 0}} \cong \frac{\sigma_p - \sigma_m}{\sigma_p + 2\sigma_m} \tag{12.11}$$

Among the parameters that influence the Clausius–Mossotti factor only the conductivity of the medium is the true variable as far as the operator is considered since the other parameters cannot be changed once the microscale entity and medium are selected. The conductivity of the medium can be adjusted by adding small amounts of ionic solutions to the medium. The influence of conductivity of the medium on Clausius–Mossotti factor can be quickly observed when operating at low frequencies as dictated by Eq. (12.11). With regard to DEP-based microfluidic devices handling cells, the operating frequencies are low and thus conductivity is a major operational parameter. It is important to mention here that there are limitations to altering the conductivity of the medium to achieve the desired Clausius–Mossotti factor. The limitation manifests through Joule heating [25]. Joule heating occurs in conductive medium and is proportional to many factors including the medium's conductivity [25]. Thus care has to be taken, while altering the conductivity of the medium, to ensure that Joule heating is avoided in DEP-based microfluidic devices; DEP-based microfluidic devices handling cells should be operated preferably at low conductivity of medium [14].

Two main conditions are needed to induce DEP force on microscale entities; nonuniform electric field ($\nabla (E \cdot E) \neq 0$); and unequal complex permittivities, i.e.,
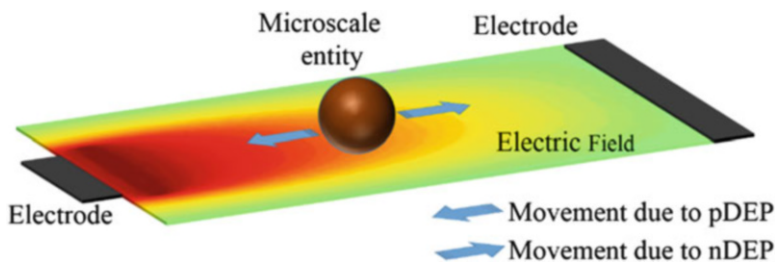
Fig. 12.1 Schematic illustration of dielectrophoresis

$Re[f_{CM}] \neq 0$. The DEP force could be positive or negative depending on the $Re[f_{CM}]$. Positive values of the real part of Clausius–Mossotti factor results in positive-DEP (pDEP) force and microscale entities will move towards high electric field gradient regions, Fig. 12.1. On the other hand, negative values of the real part of Clausius–Mossotti factor will make microscale entities move towards regions of low electric field gradient or no electric field as shown in Fig. 12.1; this is referred to as negative-DEP (nDEP). The frequency at which the change from pDEP to nDEP or vice versa occurs ($Re[f_{CM}] = 0$) is called crossover frequency.

Two primary approaches for realizing a nonuniform electrical field, for DEP, include 2D or 3D microelectrodes [14, 26] and insulator structure between electrodes [27]. 2D thin-film microelectrodes are usually deposited on the surface of the microdevice and patterned using standard photolithography process followed by wet or dry etching. Figure 12.2a shows generic example for standard 2D electrodes scheme in a microfluidic platform. 3D electrodes are less common as they require more complicated fabrication techniques such as electroplating [28]. Alternating current (AC) with frequency ranging between 1 and 10,000 kHz is usually employed to generate nonuniform electric field. Insulator-based dielectrophoresis is achieved by fabricating multiple insulator structures on the surface of microfluidic platforms to alter the electric field generated between two electrodes often placed at the inlet and outlet of the microdevice as illustrated in Fig. 12.2b.

The underlying principle of separating specific types of microscale entities from a mixture using dielectrophoresis is by assuring that the DEP force on targeted microscale entities is different from that experienced by rest of the mixture. Theoretically, two microscale entities of the same size could be separated using DEP if they have different $Re[f_{CM}]$ and two microscale entities with the same f_{CM} could be separated if they are different in size. Currently, there is enormous interest in developing DEP-based microfluidic platforms for biological and clinical applications [29]. DEP microdevices are one of the promising methods in cells assay as they are electrically manipulatable and neither require special substrates nor sheath flow [30]. DEP is used for directing, separating, and fractionating microparticles, cells, and microorganisms at small scales making the method ideal for point-of-care (POC) applications and micro total analysis systems (μTAS).
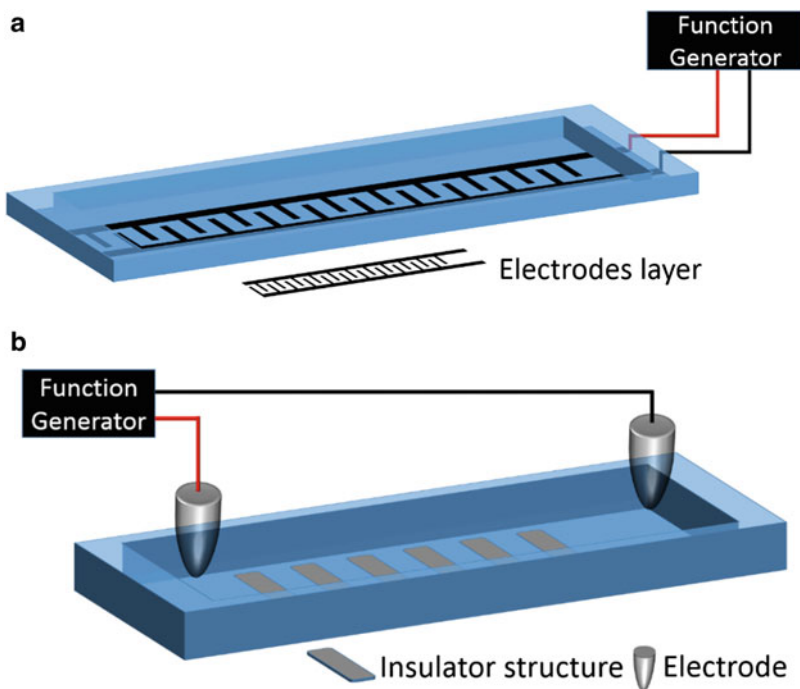
Fig. 12.2 Generic example of microfluidic device employing (**a**) 2D electrodes-based DEP and (**b**) insulator-based DEP

Several DEP-based microfluidic platforms have been investigated for separation of microscale entities. DEP separation microdevices can be categorized into two—multistep and continuous separation platforms. The multistep approach can be further subdivided into two—DEP retention and DEP migration. On the other hand, continuous separation of microscale entities can be achieved by DEP deflection and DEP field-flow-fractionation (FFF).

The DEP retention is performed by generating a strong pDEP force that is capable of trapping targeted particles by retarding their motion in the microfluidic device [31–34]. The resultant force on captured microscale entities has to come to zero in order to stop their movement. The remaining microscale entities will continue moving with the flow as illustrated in Fig. 12.3a as they experience either nDEP or weak pDEP. The drag is dominant, along the direction of flow, in case of moving microscale entities causing them to flow with the medium. DEP migration is the second multistep separation method. It is performed by gathering the targeted microscale entities using DEP force at specific locations inside the separation platform and collecting them in a subsequent step [30]. For this, the separation platform is designed to have certain regions with high gradient of electric field where particles under pDEP will be collected. Figure 12.3b illustrates the principle of DEP migration method. DEP migration technique is performed in three simple
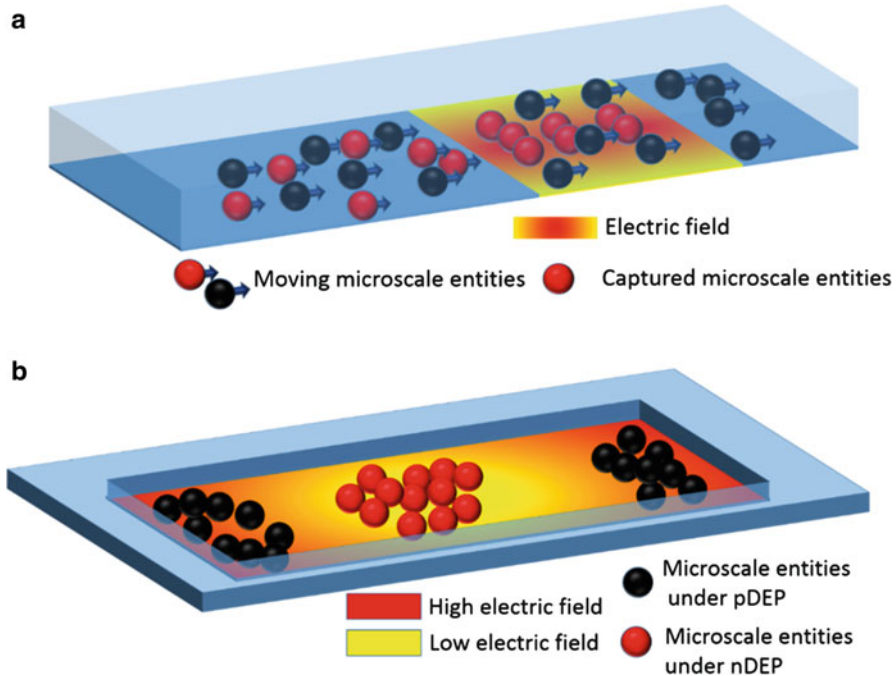
Fig. 12.3 DEP separation method (**a**) retention and (**b**) migration

steps: (1) the mixture of microscale entities is introduced into the separation device, (2) electric field is applied for a period of time to allow aggregation of microscale entities by nDEP and pDEP, and (3) targeted microscale entities are collected from designated areas. Multistep separation techniques have gained less attention than continuous methods where the separation is achieved in a single step.

Continuous DEP deflection method is performed by directing the targeted microscale entities to move towards a designated outlet while others will flow with the medium to exit from different outlet(s) and in the process achieve separation. The electric field in the separation platform is designed such that it forces the targeted microscale entities to flow from a particular outlet as illustrated in Fig. 12.4. The resultant forces on targeted microscale entities are different from the forces on other microscale entities causing them to flow in different direction. Several studies have reported DEP deflection method for separation of targeted microparticles and cells. Planar electrodes have been employed to separate platelets from diluted whole blood [35]. Cancer cells, specifically MDA231, have been successfully separated from diluted blood using interdigitated electrodes [14].

The density of microscale entities, including microparticles and cells, is usually more than the density of their medium. Therefore, microscale entities have the tendency to settle at the bottom of a microfluidic platform as a result of the difference between their density and that of the suspension medium. High gradient
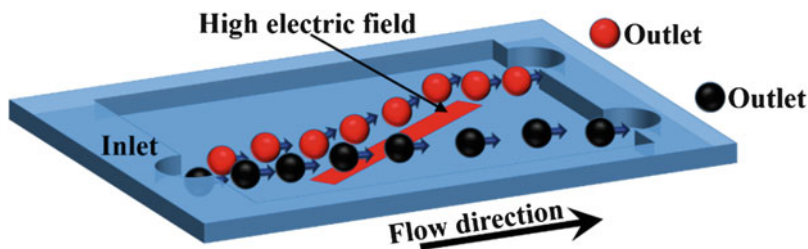
Fig. 12.4 Continuous DEP deflection separation method

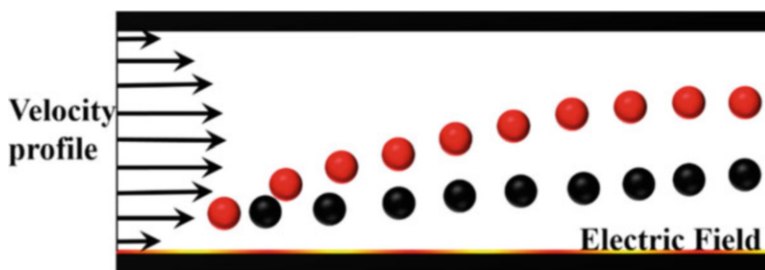**Fig. 12.5** Dielectrophoretic field-flow-fraction principle

of electric field is created at the bottom of the microfluidic separation device to create dielectrophoretic force opposite to the sedimentation force. Different microscale entities levitate to different equilibrium heights above the electric field due to the difference in forces acting on them. This phenomenon is known as dielectrophoretic field-flow-fractionation (DEP-FFF) [24, 36–38]. For a fully developed pressure driven flow, the velocity profile will have a parabolic shape. Therefore, particles at different heights will have different velocities as illustrated in Fig. 12.5. DEP-FFF has been used to separate different kinds of cells including cancer cells [39, 40], human leukocytes cells [41], and stem cells [42].

The applications of DEP microfluidic platforms go beyond the separation of microparticles and cells. DEP has been reported for separation of specific band of DNA using Au electrodes [43]. Selective transportation of cells, microparticles, and microorganism in microfluidic devices is another application of DEP force [44]. Sequential activation of single pair from an array of electrodes is used for transporting microscale entities without using pressure driven flow.

Separation of microscale entities is the main application of DEP microdevices. Other applications include sorting, moving, and switching of cells as a result of modifying their trajectories using DEP [45]. Miniaturized array of electrodes deposited at the bottom of the microchannel were used to sort cells into a small section of the stream [45, 46]. Sorting of yeast cells in a microfluidic device using planar electrodes was performed using negative dielectrophoresis [47]. Similar arrangement of electrodes was used for switching particles and cells from one medium to another and for focusing cells towards one side of the microchannel [48].

The electrodes' geometry and configuration used to create the DEP force influences the performance of the microfluidic device. Different electrode geometries have been studied for cells separation and manipulation [16, 49, 50]. These include quadrupole electrodes, interdigitated castellated electrodes, parallel electrodes, 3D irregular shape electrodes, and spiral electrodes [14, 28, 51]. Different geometries of three-dimensional (3D) electrodes were used for separation of heterogeneous mixture of microscale entities including viable and nonviable yeast cells [29, 52]. Square cross-sectional column-like electrodes were used to trap cells in dead zones using DEP retention method [26]. Isolation of white blood cells from erythrocytes mixture was achieved using parallel electrodes [53]. Other geometries of electrodes were reported for manipulation and separation of cells including annular and elliptic shape microelectrodes [54, 55]. Interdigitated castellated electrodes were widely used in DEP microfluidic devices as they offer a long trapping electric field making them an ideal geometry for DEP applications [14, 56]. Several DEP-based platforms for bio-applications have been successfully investigated and reported in the open literature. Different conductivities of medium, wide range of AC frequencies, and different configurations of microelectrodes have been used for manipulation and separation of microscale entities. New applications for DEP microfluidic systems are foreseen to appear in the near future as a result of the rapid development of DEP-based microfluidic platforms.

3   Acoustophoresis

Acoustophoresis refers to the phenomenon in which acoustic waves are employed for spatial manipulation of microscale entities including cells, microorganisms, microbeads/particles, bioparticles, and microdroplets/bubbles as schematically shown in Fig. 12.6 [57, 58]. The pressure associated with these acoustic waves is the key to enabling them in spatially manipulating microscale entities. The primary advantage of acoustophoresis is that it is a noninvasive technique that does not

Fig. 12.6 Schematic of acoustophoresis

Flow Direction

Transducer
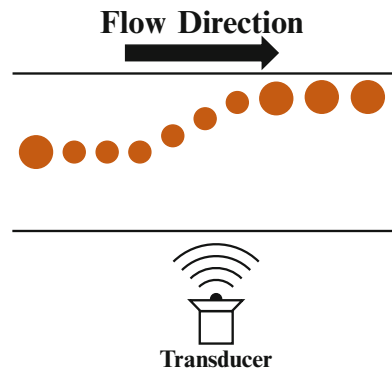
require labelling and can even be operated without sheath flow. This phenomenon has found several applications in the field of engineering such as separation, sorting, and switching of microscale entities in microfluidic devices [57, 59].

For most microfluidic applications, standing acoustic waves are employed. Standing acoustic waves are just like standing waves which are generated when two similar (same amplitude and frequency) but opposing acoustic waves are superimposed [60]. Standing acoustic wave consists of pressure nodes and antinodes [60]. Pressure node and antinode are locations on the standing acoustic wave where the pressure is maximum and minimum, respectively [60]. The distance between two consecutive pressure nodes as well as two consecutive pressure antinodes is ½ λ while the distance between each pressure node and its neighboring antinode is ¼ λ where λ is the wavelength of the standing acoustic wave [60]. The force associated with standing acoustic waves acting on a microscale entity is provided in Eq. (12.12); it is commonly referred to as primary acoustic radiation force (F_{ac}) [60, 61]. When subjected to standing acoustic waves, the microscale entity can move towards either the pressure node or antinode. The preference of microscale entities for pressure node and antinode depends on its as well as the medium's density and compressibility; these parameters are combined to form the acoustic contrast factor (ϕ_{ac}), Eq. (12.13) [57, 60]. When acoustic contrast factor is greater than zero, microscale entities will translate towards the pressure node [57, 60]. On the other hand, microscale entities will move towards the pressure antinode when acoustic contrast factor is smaller than zero [57, 60].

$$F_{ac} = \frac{8\pi^2 R_o^3}{\lambda} E_{ac} \phi_{ac} \sin\left(\frac{4\pi}{\lambda} y\right) \quad \text{where} \quad f = \frac{c}{\lambda} \tag{12.12}$$

$$\phi_{ac} = \left[\frac{5\rho_o - 2\rho_m}{2\rho_o + \rho_m} - \frac{\beta_m}{\beta_o}\right] \tag{12.13}$$

where R_o (m) is the radius of the microscale entity, E_{ac} (J m^{-3}) is the acoustic energy density of the standing acoustic wave, y (m) is the distance between the microparticle and the nearest pressure node, ρ_o (kg m^{-3}) is the density of microscale entity, ρ_m (kg m^{-3}) is the density of the medium, β_o (Pa^{-1}) is the compressibility of the microscale entity, β_m (Pa^{-1}) is the compressibility of the medium, f (Hz) is the frequency of the acoustic waves, and c (m s^{-1}) is the velocity of the acoustic wave.

It can be noticed from Eq. (12.12) that the acoustic radiation force depends on several factors including the radius and location of the microscale entity as well as the wavelength of the standing acoustic wave. With regard to the influence of radius on the force, the force increases with increase in the radius. Reduction in wavelength has a positive influence on the acoustic radiation force and this scaling bodes well for microfluidic applications [61].

Figure 12.7 schematically represents the behavior of a heterogeneous mixture (red and green) of microscale entities, in a confined space, when subjected to standing acoustic wave. It can be noticed that there is only one pressure node and two pressure antinodes along the width of the confined space. The pressure
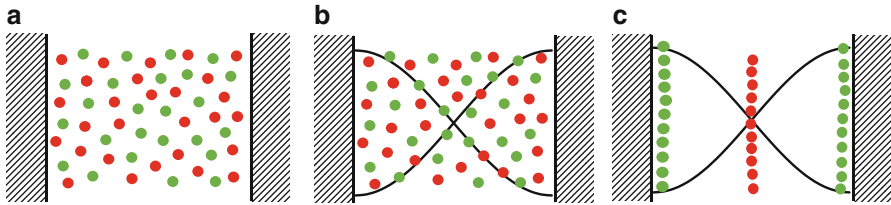
Fig. 12.7 Separation of microscale entities via standing acoustic waves—(**a**) a heterogeneous sample, (**b**) microscale entities subjected to standing acoustic waves, and (**c**) heterogeneous sample split into homogeneous samples
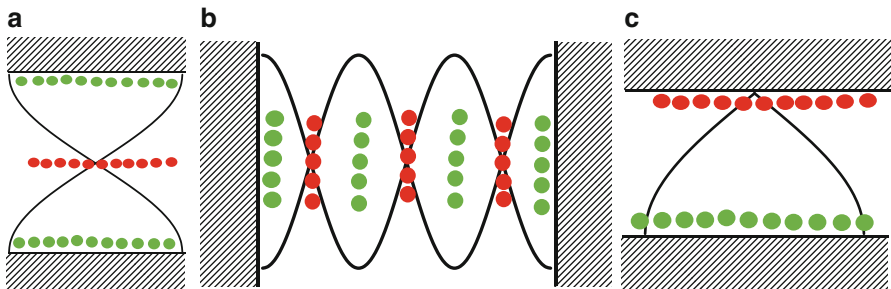

Fig. 12.8 Establishing standing acoustic wave in confined spaces—(**a**) along the height, (**b**) multiple nodes and antinodes, and (**c**) one node and antinode

node is at the center of the microchannel while the pressure antinodes occupy the sidewalls. In this case it is assumed that the acoustic contrast factor of red and green microscale entities is positive and negative, respectively. When subjected to this standing acoustic wave, the red microscale entities will move towards the center of the confined space and be settled in the vertical plane through the center of the same. On the other hand, the green microscale entities will move towards the walls of the confined space and ultimately gather in the vertical planes passing through the sidewalls. At this stage, the relevance of limiting the pressure nodes and antinodes inside the confined space can be assumed. This behavior of microscale entities in the presence of standing acoustic waves, i.e., movement towards pressure node or antinode, is the basis of several microfluidic applications for which they are currently employed [59, 62].

With regard to the orientation of standing acoustic waves, it is possible to set up standing acoustic waves (one pressure node and two pressure antinodes) along the height of the confined space as well. In this case the red microscale entities would be held on the horizontal plane passing through the center of the confined space while the green microscale entities would occupy the horizontal planes passing through the top and bottom walls of the confined space, Fig. 12.8a.

The relevance of limiting the number of pressure nodes and antinodes can be concluded from Figs. 12.7 and 12.8a; however, this need not always be the case [63–65]. For example, when there are multiple pressure nodes and antinodes inside

the confined space, the behavior would be such that each red and green microparticle will move towards the nearest pressure node and antinode, respectively; this is shown in Fig. 12.8b. It is also possible to limit the number of pressure nodes and antinodes to unity, as shown in Fig. 12.8c, in which case all the red and green microparticles would be located at the single pressure node and antinode, respectively [66].

From Figs. 12.7 and 12.8 it can be concluded that at least one of the dimensions of confined spaces, such as microchannels, should be a function of the half wavelength (½ λ) of the standing acoustic wave. For example, the width of the confined spaces shown in Fig. 12.7 and the height of the microchannel of Fig. 12.8a are ½ λ so that only one pressure node and two antinodes exist inside the microchannel. On the other hand, the height of the confined space of Fig. 12.8c is ¼ λ so as to limit the number of pressure nodes and antinodes inside the microchannel to unity. Thus in conclusion, the dimension of the confined spaces, such as microchannels, along which the standing acoustic wave is established should be a function of the wavelength of the same. With regard to microfluidic applications, standing acoustic waves established along the width of the microchannel are preferred over those along the height. This preference for the orientation of standing acoustic waves is due to the ease it provides in visualizing the interior of the microchannel as well as in extracting the processed samples downstream of the standing acoustic waves [60, 61].

Till now this section provided an overall description of the standing acoustic waves. The following part of this section will detail the practical approaches by which standing acoustic waves are established inside a microchannel [57, 61]. One of the approaches involves employing a transducer and a reflector to generate the standing acoustic wave, Fig. 12.9a [61]. In this approach the transducer generates an acoustic wave which upon striking the reflector is reflected towards the source. The original and reflected acoustic waves superimpose to generate the standing acoustic wave. Standing acoustic waves generated in this method are referred to as
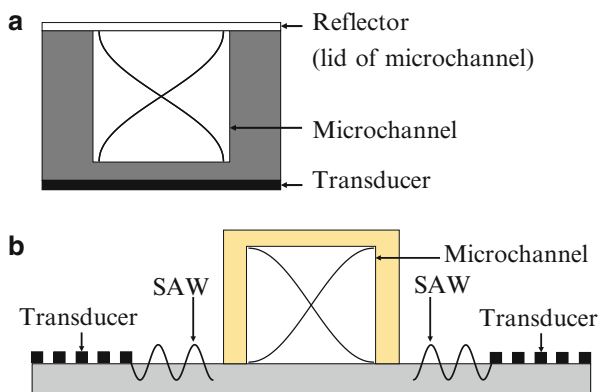

Fig. 12.9 Standing acoustic waves inside microchannels—(**a**) BAW and (**b**) SAW

Bulk **A**coustic **W**aves (BAW) [67]. In this method a piezoelectric substrate attached to the bottom of the substrate containing the microchannel acts as the transducer. Acoustophoretic microdevices of this type are also referred to as resonators [68]; acoustophoretic microdevices employing ½ λ and ¼ λ are referred to as half-wavelength and quarter-wavelength resonators, respectively [68]. In another approach two independent **S**urface **A**coustic **W**aves (SAW) are superimposed to create the standing acoustic waves, Fig. 12.9b [57, 69]. SAW are transverse sound waves traveling on the surface of solid objects including substrates; SAW decay quickly with the depth of the substrate. SAWs are generated locally, in the vicinity of the microchannel, on piezoelectric substrate, and they travel towards each other and superimpose to create the standing surface acoustic wave (SSAW) inside the microchannel. In this approach the substrate containing the microchannel is attached onto the piezoelectric substrate which holds the two transducers. Majority of acoustophoretic microdevices in recent times have been based on the second approach potentially due to the easiness with which they can be fabricated in comparison with resonators [62, 70].

The following portion of this section details few applications of standing acoustic waves. Standing acoustic waves have been used for multiple applications, in the field of microfluidics, based on unit operations such as focusing, switching, and separation [71]. All applications are detailed with respect to acoustophoretic microdevices based on SSAW as they are more widely employed. The transducer, shown in Fig. 12.9b, used for generating SAW is interdigitated transducer (IDT) electrodes [69]. The wavelength of SSAW not only determines the width of the microchannel, as mentioned earlier, but also the width and pitch of the IDT electrodes. The pitch and width, Fig. 12.10, are ¼ λ and ½ λ, respectively [69].

Focusing refers to the arranging of microscale entities randomly distributed, within the volume of the microchannel, in an orderly manner [72]. When the randomly distributed microscale entities within the volume of the microchannel are arranged onto a plane, then it is referred to as 2D focusing [72]. On the other hand, if microscale entities, randomly distributed within the volume of the microchannel, are arranged to form a single file, then it is referred to as 3D focusing [72]. Focusing is a prerequisite for most microfluidic applications as it prevents microscale entities from physically obstructing each other as well as allows unhindered exposure to
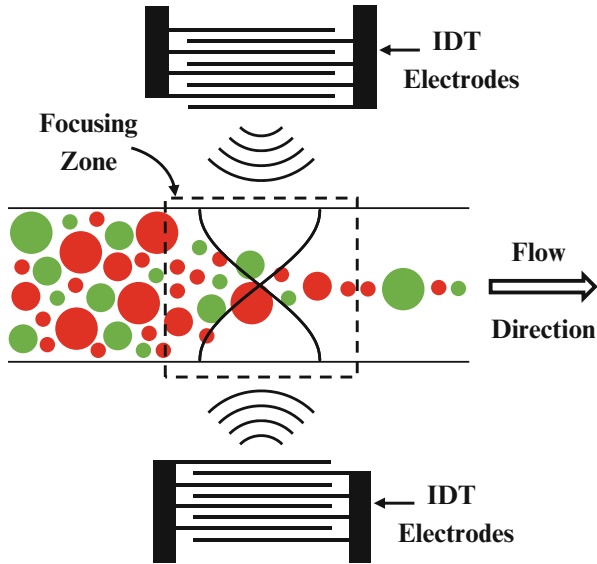
Fig. 12.10 IDT electrodes
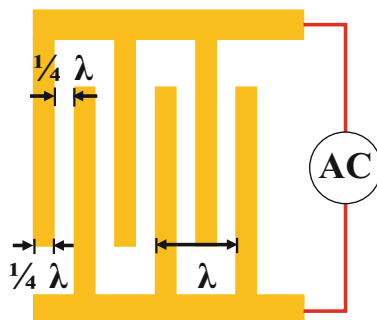
Fig. 12.11 Focusing of microscale entities via SSAW in a microchannel

the actuation force(s). SSAW has been employed to achieve 3D focusing [73]. Figure 12.11 schematically depicts focusing of acoustic contrast factor positive microscale entities in microfluidic devices. It can be noticed that the incoming microscale entities are randomly arranged while those exiting the focusing zone are arranged as a single file. The SSAW established inside the microchannel has the pressure node and antinodes at the center and sidewalls of the microchannel, respectively. Since the incoming microscale entities are acoustic contrast factor positive, they will be directed towards the pressure node at the center of the microchannel when subjected to SSAW. For focusing, the microscale entities are subjected to SSAW until all reach their steady-state location which in this case is the pressure node. As the primary acoustic radiation force, Eq. (12.6), is proportional to the size of microscale entities, the time taken to reach steady state increases with reduction in size.

Switching is another important unit operation in the field of microfluidics and researchers have demonstrated it via SSAW [74, 75]. Switching refers to the directing of incoming flow or microscale entities towards the desired downstream branch as depicted in Fig. 12.12. In convectional systems switching is carried out using mechanical valves; however, in the field of microfluidics, switching has been demonstrated without mechanical valves and for that matter switching does not hinder flow. Figure 12.12 depicts switching via SSAW. As mentioned earlier, for most applications involving SSAW, established in a microchannel, will have one pressure node and two pressure antinodes. Moreover, this pressure node would be located at the center of the microchannel while the pressure antinodes at the sidewalls. Contrary to this, if the SSAW inside the microchannel has the
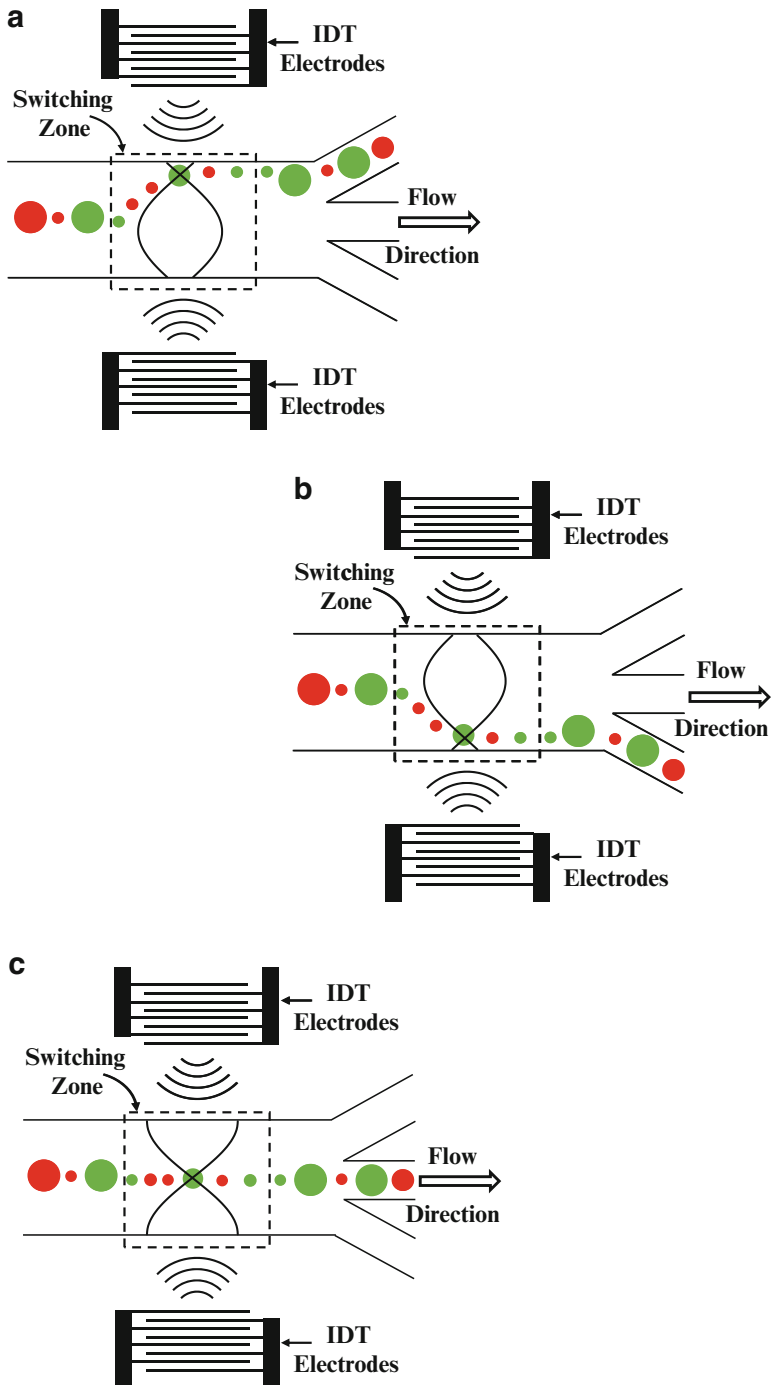
Fig. 12.12 Flow switching with SSAW; suspended microscale entities directed towards (**a**) upper, (**b**) lower, and (**c**) middle branch of microchannel

pressure node close to the upper sidewall, then the incoming stream of focused microscale entities (acoustic contrast factor positive) will be deflected towards the same, Fig. 12.12a. The fluid streamlines near the upper sidewall would then carry the microscale entities towards the upper downstream branch. In this case the pressure antinode would not be at the sidewalls; however, their locations do not matter since the microscale entities are acoustic contrast factor positive. Similarly, if the SSAW has the pressure node close to the lower sidewall, then the incoming stream of focused microscale entities will initially be deflected towards the lower sidewall by SSAW before being carried towards the lower downstream branch by fluid streamlines in the vicinity of the lower sidewall, Fig. 12.12b. If SSAW is established with the pressure node at the center, then the incoming stream of focused microscale entities would pass unaffected through the acoustic field and end up in the middle downstream branch as shown in Fig. 12.12c. In order to achieve switching, the microscale entities should be subjected to SSAW until every microscale entity reaches their steady-state location. For purposes of switching, care is required to ensure that the pressure node is at the desired location. Moreover, there should only be one pressure node inside the microchannel for obvious reasons. To the knowledge of the authors, two approaches have been devised for switching microscale entities via SSAW [74, 75]. One of the approaches is to use IDT electrodes, i.e., transducers, with varying pitch/width as this enables generating SSAW with several different wavelengths [74]. In another approach translation of the pressure node and in turn switching is achieved by introducing a phase difference between the two SAW used for generating SSAW [75].

Separation of a heterogeneous sample into homogeneous samples is a unit operation for which SSAWs are widely employed [76]. Figure 12.13 depicts the
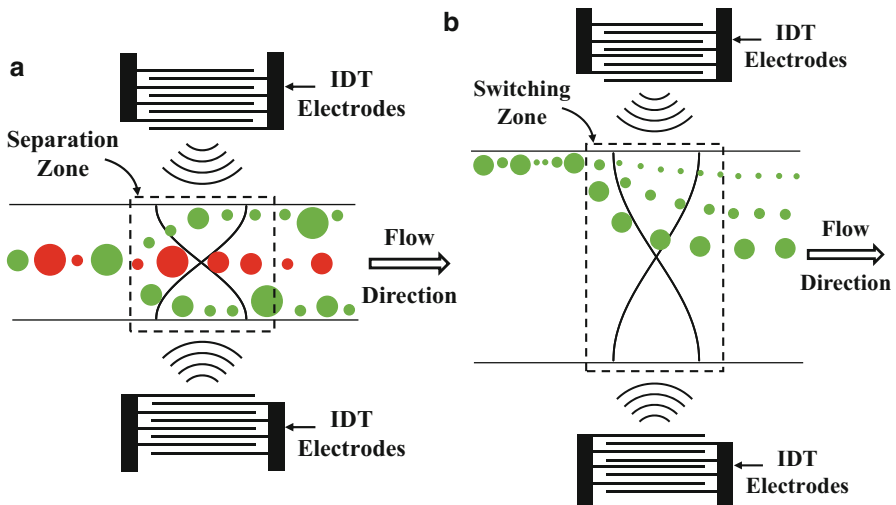

Fig. 12.13 Separation of microscale entities via SSAW—(**a**) binary and (**b**) kinematic separation

separation of microscale entities via SSAWs. Figure 12.13a represents the case where the heterogeneous sample is split into homogeneous samples based on acoustic contrast factor. The separation via SSAWs based on acoustic contrast factor is straightforward as all microscale entities can be categorized as either acoustic contrast factor positive or acoustic contrast factor negative. Thus, whenever a heterogeneous sample is subjected to SSAW in a microchannel with width equal to half the wavelength of SSAW, then the sample will be split between the pressure node and the antinode as shown in Fig. 12.13a. The microscale entities collected at the pressure node will be acoustic contrast factor positive while those at the pressure antinodes will be acoustic contrast factor negative thereby achieving the desired separation. The fluid flow in all microfluidic devices is laminar and for this matter the separated samples do not mix after exiting the separation zone. One of the characteristics of this type of separation is that each homogeneous sample contains microscale entities of different sizes; this can be a concern for certain applications thereby limiting the applicability of this approach of separation via SSAWs. This approach of separation via SSAW will result only in two homogeneous samples for that matter this separation is specifically termed as binary separation [57]. In this case, the length over which the microscale entities are subjected to SSAW has to be large as separation is achieved only when each microscale entity reaches its steady-state position, i.e., either pressure node or antinode.

In many situations it is not possible to split a heterogeneous sample of microscale entities into multiple homogeneous samples based on acoustic contrast factor. Even in such circumstances separation can still be achieved via SSAW, Fig. 12.13b [77]. In these situations, the heterogeneous sample is subjected to SSAWs for a duration which is smaller than the time required for microscale entities to reach their respective steady-state locations. During this short duration, displacement of the microscale entity depends on their size since the primary acoustic radiation force, Eq. (12.6), acting on each microscale entity is proportional to its size. Thus the biggest microscale entity would experience the maximum displacement while the smallest microscale entity will experience the minimum displacement. This difference in displacements allows for separating the heterogeneous sample into homogeneous samples based on size as shown in Fig. 12.13b. The case represented in Fig. 12.13b is based on the assumption that all incoming microscale entities are acoustic contrast factor positive; this need not always be the case and separation can be achieved even when the incoming microscale entities are acoustic contrast factor negative. For the case shown in Fig. 12.13b the microscale entities will not go past the pressure node, when subjected to SSAW, as they are acoustic contrast factor positive. This approach of separation via SSAW can be categorized as kinematic separation since the separation is completed during the transient phase of the displacement [76].

SSAW-based microfluidic devices have found several applications in the field of biomedical engineering. All microfluidics applications based on SSAW employ the abovementioned unit operations either individually or in combination. For instance, the applicability of SSAW-based focusing for creating microscale cytometer has been demonstrated [78]. Another example is that of switching of individual

droplets between five different outlets using SSAW [79]. Example of SSAW-based separation includes the microfluidic device for purposes of extracting Escherichia Coli (bacteria) in peripheral blood [80].

4   Magnetophoresis

Magnetophoresis refers to the movement of a magnetic microscale entity relative to the surrounding media, usually a nonmagnetic fluid. This relative motion is induced when the net magnetic force acting on the suspended microscale entity, referred to as magnetophoretic force, is strong enough to overcome the surface and body forces exerted by the surrounding media, Fig. 12.14. A strong but homogeneous magnetic flux cannot exert a net magnetic force on an object. Magnetophoretic force (F_{mag}) on a spherical object is mathematically stated as in Eq. (12.14) [81].

$$F_{mag} = \frac{2}{3} \frac{\pi R_o^3 (\chi_o - \chi_m)}{\mu_o} \nabla |B|^2 \tag{12.14}$$

where R_o (m) is the radius of the magnetic microscale entity, χ_o (−) is magnetic susceptibility of the microscale entity, χ_m (m mol^{-1}) is the magnetic susceptibility of the medium, μ_o (T m A^{-1}) is magnetic permeability of free space, and B (Tesla or T) is magnetic flux density.

Based on Eq. (12.14), as well as the definition of magnetophoresis, it can be concluded that in order to employ magnetophoresis for microfluidic applications it is important that the microscale entities have magnetic properties. However, most of the microscale entities handled in microfluidic devices do not have magnetic properties. Nevertheless, this issue is overcome by magnetically labeling microscale entities which involves tagging them with magnetic beads (micro-/nanoscale particles). The presence of magnetic beads immediately helps realize magnetophoresis. The approach of employing magnetophoresis for selectively separating targeted magnetically labeled cells from a heterogeneous sample can be referred to as **I**mmunomagnetic **S**eparation (IMS) and is, by large, the focus of this section.
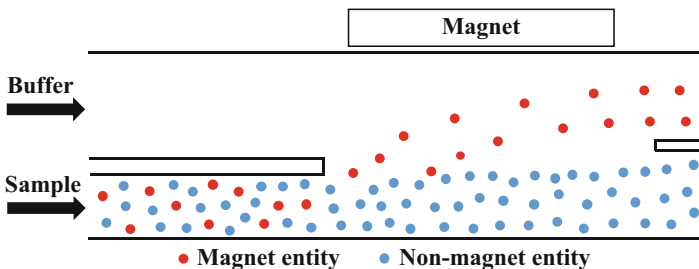

Fig. 12.14   Schematic illustration of magnetophoresis

In IMS, magnetizable micro-/nanoparticles, referred to as immunomagnetic labels, are used. The surface of these particles is synthesized from a biocompatible material and is biochemically functionalized with biomarkers, like antibodies or lectins to bind to specific receptors on the surface of a desired target (cells, pathogenic microorganisms, nucleic acids, peptide, protein, or protein complex, etc.). Since the interaction is relying on the specific affinity of the ligand on the surface of the magnetic beads, the reaction taking place between the beads and the target biological entity can be referred to as immunomagnetic labeling. The high specificity and efficiency of magnetic methods make them quite useful for purposes of separation.

Functionalized magnetic particles are commercially available from a number of vendors with sizes ranging from a few nanometers up to a few tens of microns [82–84]. The majority of these particles have a single magnetic core or dispersion of superparamagnetic nanoparticles encapsulated by a biocompatible material like polymers or silica. These beads retain the superparamagnetic nature of their core and, therefore, lose their magnetization, and re-disperse once the external magnetic source is removed. The size, magnetic content, and surface modification of these beads can be chosen depending on the application. They are usually supplied with their surface labeled with a specific antibody, antigen, or DNA strand to serve a specific application. Also, they can be supplied with rather generic surface modification including carboxyl, amino, or streptavidin groups so they can be further customized for a selective reaction. Immunomagnetic separation suffers from the inherent drawbacks associated with label-based methods. The purity of separation is directly related to the affinity with which a particular target cells can be magnetically labeled. If there are several cells that are prone to labeling by the same magnetic particle, the separation of target cells will not be as distinct as desired.

It is worth mentioning that certain magnetic labeling can be achieved without using beads; ferritin as a soluble magnetic label was used to separate lymphocytes. However, this was possible only under slow flow conditions and much stronger magnetic field and gradients [81]. For few cases, magnetic separation can be performed without labeling if the target cells are inherently susceptible to magnetization, like RBCs and magnetotactic bacteria [85–87].

In magnetophoresis-based microfluidics, magnetic force can be used for transporting, immobilizing, separating, and sorting of magnetizable as well as non-magnetizable microscale entities. The mobilizing and trapping of magnetically labeled cells are key tasks to conduct sampling, treatment, and analysis of cells. Immunomagnetic separations of a single magnetically labeled target from a multi-populated sample have been studied extensively and applied to biomedical processes such as isolating leukocytes, pathogens, and circulating tumor cells (CTCs) from the whole human blood [88–91].

Magnetophoretic microfluidic devices can operate over wide ranges of pH, ionic concentrations, and temperatures typically used in chemical and biological processing [92]. Furthermore, they have minimal, or at least, nonlethal effect on the cells [93]. In fact, the typically short magnetic exposure time diminishes the progressive harmful impacts experienced by cultured cells when subjected to long exposures [94]. Furthermore, when superparamagnetic beads (diameter

~10–100 nm) are used, their extremely small size makes them very gentle on the cells with apparently no effect on the cellular function and cell viability [95]. Negative cell enrichment, in which cells to be discarded are labeled, can eliminate the effects of magnetophoretic force related stress on target cells.

In general, the magnetic actuation of particles may involve different modes like focusing, deflecting, mobilizing, and immobilizing. The immobilization of particles is usually employed when batch-like processing is required. Deflection is usually employed with continuous free-flow processing. Mobilization, or transportation, is conducted by dragging the particle along with a carrier fluid, i.e., in pressure driven mode, or by pulling the fluid-borne particles by a directional magnetophoretic force. The magnetic separation can be classified as either trapping (capture) based or continuous-flow based.

Capture-based separation is a common method for performing batch-like particle-based processing. In the pressure driven mode, magnetically labeled particles are pumped into a microchannel and magnetically attracted towards a capture site. Depending on the application, a reagent solution may be pumped over the immobilized particles, allowing desired assays or biochemical reactions on their surfaces. The particles can then be collected/detected as they are, washed away or released for further downstream processing, Fig. 12.15 [96]. In this method, the magnetic force with the required intensity can be applied to specific locations within
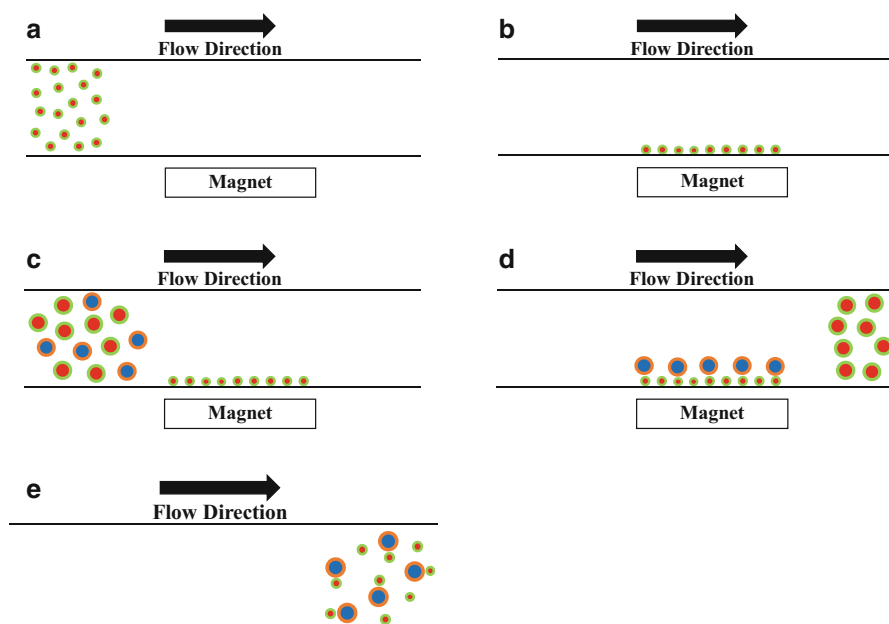
Fig. 12.15 Capture-based separation—(**a**) magnetic beads are introduced, (**b**) magnetic beads trapped by magnet, (**c**) sample containing target cells introduced, and (**d**) target cells captured, and (**e**) magnet removed to release target cells
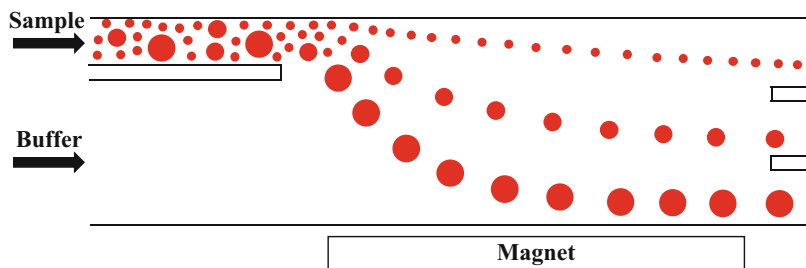
Fig. 12.16 Continuous-flow magnetophoresis

a microchannel, allowing fine control over particles trapping sites. Capture-based separation provides a simple method to perform surface processing for the particles. However, the process is not continuous and may require multiple reaction and washing steps that are time-consuming. Also, the handling capacity of these devices is limited to relatively small samples and durations due to the accumulation or irreversible entrapment that can interfere with the flow and, therefore, with further introduction of the sample. Their continuous operations are repeatedly disrupted to elude the collected particles. The setup between these batch-like consecutive steps may require changing the reagent and adjusting the pumping and other flow handling operations. After each adjustment, the system has to endure some transient time before equilibrating and attaining the steady-state capture efficiency. With such batch method, the results of an experiment/processing are available only after all steps have been completed.

In continuous-flow separation based on magnetophoresis, the magnetic microscale entities are not trapped within the microchannel by the acting magnetic force. Instead, their path gets deflected, laterally across the width of the main channel into the desire outlet, according to their dealings with the applied magnetic field acting perpendicular or at an angle to the laminar flow, Fig. 12.16. This separation technique usually necessitates designs with relatively wide and long separation chambers in order to realize significantly distinctive deflections from the main flow direction. The subpopulations of microscale entities exiting the outlets can be continuously, yet independently, collected or allowed to escape for further downstream processing. This separation mode allows the continuous processing of a sample without limits on duration or quantity. The hydrodynamics and magnetic parameters of the process can be adjusted during operation in contrast to capture-based separation in which adjustments are possible only after terminating the process. In general, multitasking can be spatially, not temporally, sequenced, thus making the integration of continuous separation with upstream and downstream processing more feasible. In many cases the magnetophoretic force is very small, especially when relying on the native magnetic properties (paramagnetic or diamagnetic) of target species. In such cases, the gradual deflection due to the magnetophoretic force (attractive or repulsive) can be easier to realize when using the continuous-flow mode. Several existing continuous-flow,
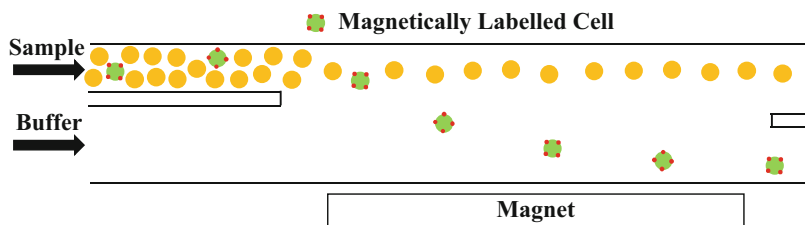
Fig. 12.17 Magnetically assisted cell sorting (MACS)

magnetophoresis-based microfluidic devices have been reviewed in literature [83]. In Fig. 12.16, the separation is based on size; however, continuous-flow separation can be based on magnetic susceptibilities as well, this can be clearly observed from Eq. (12.14).

The continuous-flow separation technique has been applied to separation of nonmagnetic microscale entities, specifically cells, as well. In this case the heterogeneous sample of cells is initially mixed with magnetic beads to magnetically label the desired cells. Afterwards the heterogeneous mixture is pumped through the continuous-flow magnetophoretic separation device during which the magnetically labeled cells would be separated from the heterogeneous sample as shown in Fig. 12.17 [97–100]. This approach of separating desired cells from a heterogeneous mixture of cells using magnetophoretic force is commonly referred to as **M**agnetically **A**ssisted **C**ell **S**orting (MACS). Both positive and negative enrichment are possible with regard to MACS. Figure 12.17 specifically shows positive cell enrichment where the desired cells, from heterogeneous sample, are labeled and separated via magnetophoresis. On the other hand, in negative cell enrichment the undesired cells are labeled and separated via magnetophoresis so that the remaining sample contains just the desired cells. Continuous-flow separation activated by magnetic force offers a simple means for performing relatively higher throughput and cell sorting. The option of real-time monitoring enables simpler tuning of the magnetic forces when necessary.

As mentioned earlier the magnetic field needs to be inhomogeneous for realizing magnetophoresis. A simple approach to induce magnetophoretic force on the magnetic particle involves employing a magnet external to the microfluidic device [98, 100–102]. Care has to be taken to ensure that the magnetic field inside the microchannel is inhomogeneous. As microfluidic devices have small substrates, it is possible to locate external magnets very close to the microchannel to potentially maintain an inhomogeneous magnet field inside the same. Nevertheless, this approach requires magnets with high flux densities. Cleanroom microfabrication techniques permit the fabrication and integration of a variety of microscale magnetic materials on microfluidic chips.

Microfabrication can be performed to realize electromagnets on the microfluidic device itself. Microfabricated circuits of current-carrying conductors patterned on the surface of a microchannel can generate field densities of several hundreds of mT

and gradients of thousands of mT/m within the volume of a microfluidic system [103–107]. Magnetic separation using micro coils has been reported [108–110]. The magnetic attributes and, hence, the strength and direction of the required magnetic force can be controlled by tuning the current flowing through a circuit or an array of closely spaced circuits. Single circuit with different wire configurations has been demonstrated to provide a controlled and localized trapping of magnetic beads [111]. Sequential switching of closely patterned circuits generates a wave of moving field maxima that can be utilized to transport particles along its direction [112]. Such patterned structures can capture particles from a solution, immobilize them in a specified position, and transport them along specified routes. The patterned structures can be coated for biocompatibility as well for achieving surface functionalities vital for some biotechnological and chemical applications. Magnetic field density and gradients generated by microfabricated current-carrying circuits can be further improved by integrating them with miniaturized ferromagnetic pillars and/or attaching to magnetic backside plates [111]. Also, external electromagnetic fields can be concentrated near the capture site by attaching bulky electromagnets to soft magnetic microtips [113].

Another effective approach towards developing inhomogeneous magnetic field involves integrating ferromagnetic microstructures within the microchannel and subjecting the entire microfluidic device to an external magnetic field. The presence of the ferromagnetic microstructure makes the external magnetic field inhomogeneous in its vicinity. These elements, like permalloy and nickel, redirect the external magnetic flux lines through their highly permeable structures and therefore induce localized gradient concentrations. In that sense, these structures can be referred to as **H**igh-**G**radient **M**agnetic field **C**oncentrators (HGMC). With this approach the constraint on the distance between the microfluidic device and the external magnet is less stringent than when external magnets are employed. Even this approach requires magnets with high flux densities and the capture efficiency is not necessarily scalable with number of embedded elements [114]. The passive magnetic actuation attained by integrating these soft elements to a microfluidic system removes many Joule heating problems that usually arise when using electric conductors. Most often an array of thin film ferromagnetic microstructures is realized on (or near) the bottom surface of the microchannel [115–117]. For continuous-flow separation, ferromagnetic stripes can be patterned on the microchannel wall to make an angle with respect to the sample flow. These externally magnetized strips create a series of localized high gradients that can deflect the labeled cells along their length towards a spatially addressable outlet [88, 118]

The magnetic separation can be greatly enhanced by using flow-invasive soft-magnetic elements. In a recently proposed design, these elements intersect the carrier fluid and are arranged transverse to the flow in a vertical stair step configuration that spans the height of the flow channel [119]. Computational analysis indicates that such an arrangement enables substantially higher capture efficiency than that enabled by conventional systems where the magnetic elements are embedded in the walls of the flow chamber and do not intersect the flow.

Table 12.1 Quick comparison of dielectrophoresis, acoustophoresis, and magnetophoresis

Method	Dielectrophoresis	Acoustophoresis	Magnetophoresis
Field	Nonuniform electric field	Acoustic pressure field	Nonuniform magnetic field
Separation criteria	Size and Clausius–Mossotti factor (permittivities and conductivities)	Size and acoustic contrast factor (densities and compressibilities)	Size and susceptibilities
Equation	Equations (12.7)–(12.9)	Equations (12.12) and (12.13)	Equation (12.14)
Characteristics	Noninvasive, label-free, electrically controllable, susceptible to joule heating, needs adjustment of medium conductivity, commonly fabricated using techniques such as photolithography, thin film deposition, wet etching, and bonding	Noninvasive, label-free, electrically controllable, requires specialized wafers, susceptible to heating, commonly fabricated using techniques such as photolithography, thin film deposition, wet etching, and bonding	Noninvasive, requires labeling (for microscale entities without magnetic properties), electrically controllable when employing electromagnets, commonly fabricated using techniques such as photolithography, thin film deposition, wet etching, electroplating, and bonding

5 Conclusion

This chapter provides an overview of actuation mechanisms such as acoustophoresis, dielectrophoresis, and magnetophoresis. The chapter details existing approaches, involving these actuation mechanisms, for executing unit operations such as focusing, switching, and separation. Table 12.1 provides a comparison of the different actuation mechanisms dealt with in this chapter. These actuation mechanisms are suitable for a wide range of biomedical application; however, the preference for particular actuation mechanism depends on the application as well as the available resources.

References

1. Terry SC, Jerman JH, Angell JB (1979) A gas chromatographic air analyzer fabricated on a silicon wafer. IEEE Trans Electron Devices 26:1880–1886
2. Nguyen N-T, Wereley ST (2002) Fundamentals and applications of microfluidics. Artech House, Norwood
3. Folch A (2012) Introduction to BioMEMS. CRC, Boca Raton
4. Saliterman S (2006) Fundamentals of BioMEMS and medical microdevices, vol 153. SPIE Press, Bellingham

5. Gervais T, Jensen KF (2006) Mass transport and surface reactions in microfluidic systems. Chem Eng Sci 61:1102–1121
6. Jensen KF (1999) Microchemical systems: status, challenges, and opportunities. AIChE J 45:2051
7. Kutter JP, Fintschenko Y (2005) Separation methods in microanalytical systems. CRC, Boca Raton
8. Deen WM (1998) Analysis of transport phenomena, vol 3, Topics in chemical engineering. Oxford University Press, New York
9. Stone HA, Stroock AD, Ajdari A (2004) Engineering flows in small devices: microfluidics toward a lab-on-a-chip. Annu Rev Fluid Mech 36:381–411
10. White F (2006) Viscous fluid flow, McGraw-Hill series in mechanical engineering. McGraw-Hill, New York
11. Sackmann EK, Fulton AL, Beebe DJ (2014) The present and future role of microfluidics in biomedical research. Nature 507:181–189
12. Ziaie B, Baldi A, Lei M, Gu Y, Siegel RA (2004) Hard and soft micromachining for BioMEMS: review of techniques and examples of applications in microfluidics and drug delivery. Adv Drug Deliv Rev 56:145–172
13. Pohl HA (1978) Dielectrophoresis: the behavior of neutral matter in nonuniform electric fields, vol 80. Cambridge University Press, Cambridge
14. Alazzam A, Stiharu I, Bhat R, Meguerditchian AN (2011) Interdigitated comb-like electrodes for continuous separation of malignant cells from blood using dielectrophoresis. Electrophoresis 32:1327–1336
15. Chen D, Du H, Li W (2007) Bioparticle separation and manipulation using dielectrophoresis. Sens Actuators A Phys 133:329–334
16. Alazzam A, Roman D, Nerguizian V, Stiharu I, Bhat R (2010) Analytical formulation of electric field and dielectrophoretic force for moving dielectrophoresis using Fourier series. Microfluid Nanofluid 9:1115–1124
17. Guo X, Zhu R (2015) Controllably moving individual living cell in an array by modulating signal phase difference based on dielectrophoresis. Biosens Bioelectron 68:529–535
18. Zhang J, Yan S, Alici G, Nguyen N-T, Di Carlo D, Li W (2014) Real-time control of inertial focusing in microfluidics using dielectrophoresis (DEP). RSC Adv 4:62076–62085
19. Zhu H, Lin X, Su Y, Dong H, Wu J (2015) Screen-printed microfluidic dielectrophoresis chip for cell separation. Biosens Bioelectron 63:371–378
20. Huang Y, Joo S, Duhon M, Heller M, Wallace B, Xu X (2002) Dielectrophoretic cell separation and gene expression profiling on microelectronic chip arrays. Anal Chem 74:3362–3371
21. Pethig R, Markx GH (1997) Applications of dielectrophoresis in biotechnology. Trends Biotechnol 15:426–432
22. Huang Y, Holzel R, Pethig R, Wang X-B (1992) Differences in the AC electrodynamics of viable and non-viable yeast cells determined through combined dielectrophoresis and electrorotation studies. Phys Med Biol 37:1499
23. Huang Y, Yang J, Wang X-B, Becker FF, Gascoyne PR (1999) The removal of human breast cancer cells from hematopoietic CD34+ stem cells by dielectrophoretic field-flow-fractionation. J Hematother Stem Cell Res 8:481–490
24. Mathew B, Alazzam A, Abutayeh M, Gawanmeh A, Khashan S (2015) Modeling the trajectory of microparticles subjected to dielectrophoresis in a microfluidic device for field flow fractionation. Chem Eng Sci 138:266–280
25. Ramos A, Morgan H, Green NG, Castellanos A (1998) Ac electrokinetics: a review of forces in microelectrode structures. J Phys D Appl Phys 31:2338
26. Iliescu C, Yu L, Tay FE, Chen B (2008) Bidirectional field-flow particle separation method in a dielectrophoretic chip with 3D electrodes. Sens Actuators B 129:491–496
27. LaLonde A, Romero-Creel MF, Lapizco-Encinas BH (2015) Assessment of cell viability after manipulation with insulator-based dielectrophoresis. Electrophoresis 36(13):1479–1484

28. Tay FE, Yu L, Pang AJ, Iliescu C (2007) Electrical and thermal characterization of a dielectrophoretic chip with 3D electrodes for cells manipulation. Electrochim Acta 52:2862–2868

29. Srinivasan V, Pamula VK, Fair RB (2004) An integrated digital microfluidic lab-on-a-chip for clinical diagnostics on human physiological fluids. Lab Chip 4:310–315

30. Gascoyne PR, Vykoukal J (2002) Particle separation by dielectrophoresis. Electrophoresis 23:1973

31. Becker FF, Wang X-B, Huang Y, Pethig R, Vykoukal J, Gascoyne P (1995) Separation of human breast cancer cells from blood by differential dielectric affinity. Proc Natl Acad Sci U S A 92:860–864

32. Broche LM, Bhadal N, Lewis MP, Porter S, Hughes MP, Labeed FH (2007) Early detection of oral cancer—is dielectrophoresis the answer? Oral Oncol 43:199–203

33. Cheng J, Sheldon EL, Wu L, Heller MJ, O'Connell JP (1998) Isolation of cultured cervical carcinoma cells mixed with peripheral blood cells on a bioelectronic chip. Anal Chem 70:2321–2326

34. Gascoyne PR, Wang X-B, Huang Y, Becker FF (1997) Dielectrophoretic separation of cancer cells from blood. IEEE Trans Ind Appl 33:670–678

35. Pommer MS, Zhang Y, Keerthi N, Chen D, Thomson JA, Meinhart CD, Soh HT (2008) Dielectrophoretic separation of platelets from diluted whole blood in microfluidic channels. Electrophoresis 29:1213–1218

36. Camarda M, Scalese S, La Magna A (2015) Analysis of the role of the particle–wall interaction on the separation efficiencies of field flow fractionation dielectrophoretic devices. Electrophoresis 36(13):1396–1404

37. Mathew B, Alazzam A, Khashan S, El-Khasawneh B (2016), Path of microparticles in a microfluidic device employing dielectrophoresis for hyperlayer field-flow fractionation. Microsystem Technologies, pp 1–12

38. Wang X-B, Vykoukal J, Becker FF, Gascoyne PR (1998) Separation of polystyrene microbeads using dielectrophoretic/gravitational field-flow-fractionation. Biophys J 74:2689–2701

39. Huang Y, Wang X-B, Becker FF, Gascoyne P (1997) Introducing dielectrophoresis as a new force field for field-flow fractionation. Biophys J 73:1118–1129

40. Yang J, Huang Y, Wang X-B, Becker FF, Gascoyne PR (1999) Cell separation on microfabricated electrodes using dielectrophoretic/gravitational field-flow fractionation. Anal Chem 71:911–918

41. Yang J, Huang Y, Wang X-B, Becker FF, Gascoyne PR (2000) Differential analysis of human leukocytes by dielectrophoretic field-flow-fractionation. Biophys J 78:2680–2689

42. Vykoukal J, Vykoukal DM, Freyberg S, Alt EU, Gascoyne PR (2008) Enrichment of putative stem cells from adipose tissue using dielectrophoretic field-flow fractionation. Lab Chip 8:1386–1393

43. Petersen E et al (2007) DNA migration and separation on surfaces with a microscale dielectrophoretic trap array. Phys Rev Lett 98:088102

44. Kua CH, Lam YC, Rodriguez I, Yang C, Youcef-Toumi K (2007) Dynamic cell fractionation and transportation using moving dielectrophoresis. Anal Chem 79:6975–6987

45. Fiedler S, Shirley SG, Schnelle T, Fuhr G (1998) Dielectrophoretic sorting of particles and cells in a microsystem. Anal Chem 70:1909–1915

46. Taff BM, Voldman J (2005) A scalable addressable positive-dielectrophoretic cell-sorting array. Anal Chem 77:7976–7983

47. Christensen TB, Pedersen CM, Bang DD, Wolff A (2007) Sample preparation by cell guiding using negative dielectrophoresis. Microelectron Eng 84:1690–1693

48. Tornay R et al (2008) Dielectrophoresis-based particle exchanger for the manipulation and surface functionalization of particles. Lab Chip 8:267–273

49. Nerguizian V, Alazzam A, Roman D, Stiharu I, Burnier M (2012) Analytical solutions and validation of electric field and dielectrophoretic force in a bio-microfluidic channel. Electrophoresis 33:426–435

50. Nieuwenhuis JH, Vellekoop MJ (2004) Simulation study of dielectrophoretic particle sorters. Sens Actuators B 103:331–338
51. Wang X-B, Huang Y, Wang X, Becker FF, Gascoyne P (1997) Dielectrophoretic manipulation of cells with spiral electrodes. Biophys J 72:1887
52. Yu L, Iliescu C, Xu G, Tay FE (2007) Sequential field-flow cell separation method in a dielectrophoretic chip with 3-D electrodes. J Microelectromech Syst 16:1120–1129
53. Borgatti M et al (2005) Separation of white blood cells from erythrocytes on a dielectrophoresis (DEP) based 'Lab-on-a-chip' device. Int J Mol Med 15:913–920
54. Bocchi M, Lombardini M, Faenza A, Rambelli L, Giulianelli L, Pecorari N, Guerrieri R (2009) Dielectrophoretic trapping in microwells for manipulation of single cells and small aggregates of particles. Biosens Bioelectron 24:1177–1183
55. Yu C, Vykoukal J, Vykoukal DM, Schwartz J, Shi L, Gascoyne PR (2005) A three-dimensional dielectrophoretic particle focusing channel for microcytometry applications. J Microelectromech Syst 14:480–487
56. Li H, Bashir R (2002) Dielectrophoretic separation and manipulation of live and heat-treated cells of Listeria on microfabricated devices with interdigitated electrodes. Sens Actuators B 86:215–221
57. Lenshof A, Laurell T (2012) Acoustophoresis. In: Bhushan B (ed) Encyclopedia of nanotechnology. Springer, Berlin, pp 45–50
58. Sajeesh P, Sen AK (2014) Particle separation and sorting in microfluidic devices: a review. Microfluid Nanofluid 17:1–52
59. Qiu Y et al (2014) Acoustic devices for particle and cell manipulation and sensing. Sensors 14:14806–14838
60. Leong T, Johansson L, Juliano P, McArthur SL, Manasseh R (2013) Ultrasonic separation of particulate fluids in small and large scale systems: a review. Ind Eng Chem Res 52:16555–16576
61. Lenshof A, Evander M, Laurell T, Nilsson J (2012) Acoustofluidics 5: building microfluidic acoustic resonators. Lab Chip 12:684–695
62. Lin S-CS, Mao X, Huang TJ (2012) Surface acoustic wave (SAW) acoustophoresis: now and beyond. Lab Chip 12:2766–2770
63. Nilsson A, Petersson F, Jönsson H, Laurell T (2004) Acoustic control of suspended particles in micro fluidic chips. Lab Chip 4:131–135
64. Piyasena ME, Austin Suthanthiraraj PP, Applegate RW Jr, Goumas AM, Woods TA, López GP, Graves SW (2012) Multinode acoustic focusing for parallel flow cytometry. Anal Chem 84:1831–1839
65. Suthanthiraraj PPA, Piyasena ME, Woods TA, Naivar MA, López GP, Graves SW (2012) One-dimensional acoustic standing waves in rectangular channels for flow cytometry. Methods 57:259–271
66. Kuznetsova LA, Coakley WT (2004) Microparticle concentration in short path length ultrasonic resonators: roles of radiation pressure and acoustic streaming. J Acoust Soc Am 116:1956–1966
67. Shields CW IV, Reyes CD, López GP (2015) Microfluidic cell sorting: a review of the advances in the separation of cells from debulking to rare cell isolation. Lab Chip 15:1230–1249
68. Glynne-Jones P, Boltryk RJ, Hill M (2012) Acoustofluidics 9: modelling and applications of planar resonant devices for acoustic particle manipulation. Lab Chip 12:1417–1426
69. Gedge M, Hill M (2012) Acoustofluidics 17: theory and applications of surface acoustic wave devices for particle manipulation. Lab Chip 12:2998–3007
70. Shi J, Mao X, Ahmed D, Colletti A, Huang TJ (2008) Focusing microparticles in a microfluidic channel with standing surface acoustic waves (SSAW). Lab Chip 8:221–223
71. Mark D, Haeberle S, Roth G, von Stetten F, Zengerle R (2010) Microfluidic lab-on-a-chip platforms: requirements, characteristics and applications. Chem Soc Rev 39:1153–1182
72. Xuan X, Zhu J, Church C (2010) Particle focusing in microfluidic devices. Microfluid Nanofluid 9:1–16

73. Shi J, Yazdi S, Lin S-CS, Ding X, Chiang I-K, Sharp K, Huang TJ (2011) Three-dimensional continuous particle focusing in a microfluidic channel via standing surface acoustic waves (SSAW). Lab Chip 11:2319–2324

74. Ding X et al (2012) Standing surface acoustic wave (SSAW) based multichannel cell sorting. Lab Chip 12:4228–4231

75. Jo MC, Guldiken R (2014) Particle manipulation by phase-shifting of surface acoustic waves. Sens Actuators A Phys 207:39–42

76. Gossett DR et al (2010) Label-free cell separation and sorting in microfluidic systems. Anal Bioanal Chem 397:3249–3267

77. Shi J, Huang H, Stratton Z, Huang Y, Huang TJ (2009) Continuous particle separation in a microfluidic channel via standing surface acoustic waves (SSAW). Lab Chip 9:3354–3359

78. Chen Y et al (2014) Standing surface acoustic wave (SSAW)-based microfluidic cytometer. Lab Chip 14:916–923

79. Li S et al (2013) An on-chip, multichannel droplet sorter using standing surface acoustic waves. Anal Chem 85:5468–5474

80. Ai Y, Sanders CK, Marrone BL (2013) Separation of Escherichia coli bacteria from peripheral blood mononuclear cells using standing surface acoustic waves. Anal Chem 85:9126–9134

81. Zborowski M, Fuh CB, Green R, Sun L, Chalmers JJ (1995) Analytical magnetapheresis of ferritin-labeled lymphocytes. Anal Chem 67:3702–3712

82. Gijs MA (2004) Magnetic bead handling on-chip: new opportunities for analytical applications. Microfluid Nanofluid 1:22–40

83. Pamme N (2006) Magnetism and microfluidics. Lab Chip 6:24–38

84. Pankhurst QA, Connolly J, Jones S, Dobson J (2003) Applications of magnetic nanoparticles in biomedicine. J Phys D Appl Phys 36:R167

85. Han K-H, Frazier AB (2006) Paramagnetic capture mode magnetophoretic microseparator for high efficiency blood cell separations. Lab Chip 6:265–273

86. Jung J, Han K-H (2009) Lateral-driven continuous magnetophoretic microseparator for separating blood cells based on their native magnetic properties. In: Solid-state sensors, actuators and microsystems conference, 2009. TRANSDUCERS 2009. International, 2009. IEEE, pp 620–623

87. Schüler D, Frankel RB (1999) Bacterial magnetosomes: microbiology, biomineralization and biotechnological applications. Appl Microbiol Biotechnol 52:464–473

88. Inglis DW, Riehn R, Austin R, Sturm J (2004) Continuous microfluidic immunomagnetic cell separation. Appl Phys Lett 85:5093–5095

89. Inglis DW, Riehn R, Sturm JC, Austin RH (2006) Microfluidic high gradient magnetic cell separation. J Appl Phys 99:08K101

90. Saias L, Saliba A-E, Viovy J-L, Pierga J-Y, Vielh P, Farace F (2009) Microfluidic magnetic cell sorting system for cancer diagnosis. La Houille Blanche, pp 105–111

91. Yung CW, Fiering J, Mueller AJ, Ingber DE (2009) Micromagnetic–microfluidic blood cleansing device. Lab Chip 9:1171–1177

92. Tsutsui H, Ho C-M (2009) Cell separation by non-inertial force fields in microfluidic systems. Mech Res Commun 36:92–103

93. Šafařık I, Šafařıková M (1999) Use of magnetic techniques for the isolation of cells. J Chromatogr B Biomed Sci Appl 722:33–53

94. Dini L, Abbro L (2005) Bioeffects of moderate-intensity static magnetic fields on cell cultures. Micron 36:195–217

95. Yi C, Li C-W, Ji S, Yang M (2006) Microfluidics technology for manipulation and analysis of biological cells. Anal Chim Acta 560:1–23

96. Choi J-W, Ahn CH, Bhansali S, Henderson HT (2000) A new magnetic bead-based, filterless bio-separator with planar electromagnet surfaces for integrated bio-detection systems. Sens Actuators B 68:34–39

97. Pamme N, Manz A (2004) On-chip free-flow magnetophoresis: continuous flow separation of magnetic particles and agglomerates. Anal Chem 76:7250–7256

98. Pamme N, Wilhelm C (2006) Continuous sorting of magnetic cells via on-chip free-flow magnetophoresis. Lab Chip 6:974–980

99. Tarn MD, Peyman SA, Robert D, Iles A, Wilhelm C, Pamme N (2009) The importance of particle type selection and temperature control for on-chip free-flow magnetophoresis. J Magn Magn Mater 321:4115–4122

100. Tsai SS, Griffiths IM, Stone HA (2011) Microfluidic immunomagnetic multi-target sorting—a model for controlling deflection of paramagnetic beads. Lab Chip 11:2577–2582

101. Hayes MA, Polson NA, Phayre AN, Garcia AA (2001) Flow-based microimmunoassay. Anal Chem 73:5896–5902

102. Schneider T, Karl S, Moore LR, Chalmers JJ, Williams PS, Zborowski M (2010) Sequential CD34 cell fractionation by magnetophoresis in a magnetic dipole flow sorter. Analyst 135: 62–70

103. Deng T, Prentiss M, Whitesides GM (2002) Fabrication of magnetic microfiltration systems using soft lithography. Appl Phys Lett 80:461–463

104. Khashan SA, Elnajjar E, Haik Y (2011) CFD simulation of the magnetophoretic separation in a microchannel. J Magn Magn Mater 323:2960–2967

105. Khashan SA, Elnajjar E, Haik Y (2011) Numerical simulation of the continuous biomagnetic separation in a two-dimensional channel. Int J Multiphase Flow 37:947–955

106. Khashan SA, Furlani EP (2012) Effects of particle–fluid coupling on particle transport and capture in a magnetophoretic microsystem. Microfluid Nanofluid 12:565–580

107. Ng JMK, Gitlin I, Stroock AD, Whitesides GM (2002) Components for integrated poly (dimethylsiloxane) microfluidic systems. Electrophoresis 23:3461–3473

108. Choi J-W et al (2001) Development and characterization of microfluidic devices and systems for magnetic bead-based biochemical detection. Biomed Microdevices 3:191–200

109. Lee C, Lee H, Westervelt R (2001) Microelectromagnets for the control of magnetic nanoparticles. Appl Phys Lett 79:3308–3310

110. Ramadan Q, Gijs MA (2012) Microfluidic applications of functionalized magnetic particles for environmental analysis: focus on waterborne pathogen detection. Microfluid Nanofluid 13:529–542

111. Ramadan Q, Samper V, Poenar DP, Yu C (2006) An integrated microfluidic platform for magnetic microbeads separation and confinement. Biosens Bioelectron 21:1693–1702

112. Deng T, Whitesides GM, Radhakrishnan M, Zabow G, Prentiss M (2001) Manipulation of magnetic microbeads in suspension using micromagnetic systems fabricated with soft lithography. Appl Phys Lett 78:1775–1777

113. Afshar R, Moser Y, Lehnert T, Gijs M (2011) Three-dimensional magnetic focusing of superparamagnetic beads for on-chip agglutination assays. Anal Chem 83:1022–1029

114. Khashan SA, Furlani EP (2014) Scalability analysis of magnetic bead separation in a microchannel with an array of soft magnetic elements in a uniform magnetic field. Sep Purif Technol 125:311–318

115. Bu M, Christensen TB, Smistrup K, Wolff A, Hansen MF (2008) Characterization of a microfluidic magnetic bead separator for high-throughput applications. Sens Actuators A Phys 145:430–436

116. Khashan SA, Haik Y, Elnajjar E (2012) CFD simulation for biomagnetic separation involving dilute suspensions. Can J Chem Eng 90:1450–1456

117. Xia N, Hunt TP, Mayers BT, Alsberg E, Whitesides GM, Westervelt RM, Ingber DE (2006) Combined microfluidic-micromagnetic separation of living cells in continuous flow. Biomed Microdevices 8:299–308

118. Adams JD, Kim U, Soh HT (2008) Multitarget magnetic activated cell sorter. Proc Natl Acad Sci U S A 105:18165–18170

119. Khashan S, Alazzam A, Furlani E (2014) Computational analysis of enhanced magnetic bioseparation in microfluidic systems with flow-invasive magnetic elements. Sci Rep 4:5299

Chapter 13
Recent Advances in Mechatronics Devices: Screening and Rehabilitation Devices for Autism Spectrum Disorder

Hadi Moradi and Iman Mohammad-Rezazadeh

Abstract Autism spectrum disorder (ASD) is one of the most common, with the rate of 1 out of 68 children in the USA, neurodevelopmental disorders and its core symptoms comprise of deficits in social communication, and restricted and repetitive behaviors. Early and accurate screening and intervention are crucial factors in enhancing the quality of life and functional independency of individuals with ASD. Much research has been performed to design and develop mechatronics technologies to strengthen the performance of the screening and rehabilitation procedures for individuals with ASD. In this chapter, we review some of the currently developed devices in this area and discuss their advantages and limitations, as well as potential technologies which can be employed in such devices in the future.

1 Introduction

Autism is a neurodevelopmental disorder characterized by severe impairments in social communication and unusual, restricted, or repetitive behaviors [1]. Based on the severity of autism symptoms, a child may fall along a "spectrum" of this disorder which is called Autism Spectrum Disorder or ASD for short [1]. There have been many studies to determine the contributing causes for ASD such as genetics, environment, and maternal age [2]. In addition, many greater attempts have been

H. Moradi (✉)
School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

Intelligent Systems Research Institute, SKKU, Suwon, South Korea
e-mail: moradih@ut.ac.ir

I. Mohammad-Rezazadeh
Semel Institute for Neuroscience and Human Behavior, UCLA David Geffen School of Medicine, Los Angeles, CA, 90095 USA

done to increase the accuracy of screening, treatment monitoring process, as well as alternative therapeutic options for ASD [3].

Traditional approaches in screening ASD, which bring potentially affected children to more structured clinical assessments, rely on questionnaires which usually only assess parent repots. Normally, the questionnaires are used after parents are alerted to their children's abnormal behavior or delayed developmental progress. Unfortunately, parents usually do not notice the delayed developmental progress early enough since they are not experts in child neurodevelopment and infant developmental milestones. Furthermore, although there is no age limit for the onset of ASD symptoms, these symptoms may not become apparent until children are being demanded of skills beyond their abilities, such as the skill demands available in school environments [4]. It should also be mentioned that in many countries, there is neither general public awareness about ASD nor widespread medical services available to provide early screening to infants and children [5]. As a result, children with ASD may only have their first screening after 2 years of age, with therapy sessions beginning afterwards. This is the main reason to develop devices that can assist caregivers and parents to facilitate an early screening process.

On the other hand, after diagnosing a child with autism, the next step is to provide extensive therapy sessions to reduce the impairments of the disorder. The main therapy approach in the past half a century has been occupational therapy in which a therapist works with individual or groups of children with ASD. In many cases, more than one therapist is needed to work with children due to difficulty in controlling children with ASD. There are many issues with using only this type of intervention such as: (1) high cost of therapy for the families and the government [6]; (2) lack of widely accessible professional therapists; (3) limited available time provided by therapists compared to the extensive time and attention required by the children with ASD; and (4) lack of time, knowledge, and patience in parents of children with ASD [7]. The involvement of parents in post-therapy sessions is necessary to make the professional therapy more effective and its effect last longer. Thus, this creates a strong motivation to develop devices to assist therapists and families in the therapy process of children with ASD.

A review on the trends toward developing systems for ASD screening and rehabilitation [3] shows that about 64 % of the published papers on ASD screening and rehabilitation have been primarily focused on developing software packages for rehabilitation, while the rest of the papers centered on hardware developments. The study shows that about 88 % of the hardware-related papers report the usage of robots for ASD screening and rehabilitation and only 12 % report the development of devices dedicated for this purpose. This trend is seemingly logical since the development of software-based packages is much easier than hardware-based systems. Furthermore, in the hardware-based category, using available robots is easier than developing new and dedicated devices.

However, recent advancements in mechatronic devices provide better opportunities to develop dedicated devices suitable for ASD screening and rehabilitation. For instance, the new era of IoT (Internet of Things) ignited the introduction of small devices such as the numerous tiny and small mechatronic systems developed based

on ESP8266 chipset [8], which includes an MCU (Micro Controller Unit) for all input and output operations and Wi-Fi for communication with other devices. Thus, this type of device, in conjunction with other devices/sensors (such as MEMS-based accelerometers) can be used to collect and transfer data to a centralized server for ASD screening procedures. This approach leads to collection of data from large group of subjects and consequently normative databases for more accurate and sensitive screening procedure.

Another field, which has already stepped into its second decade of life, is sensor networks in which sensors create a network to handle a task in a physically distributed environment. These mechatronic devices allow collecting, processing, and sharing behavioral and physiological data in real-time for better screening and rehabilitation. This would allow several sensors to be placed around a child's living environment to collect data and process them for screening and therapy purposes.

In this chapter, we first review a few mechatronic devices designed and used in ASD screening. Then, we discuss a few devices used for therapeutic purposes. Finally, we summarize the findings and propose possible future directions on assistive and therapeutic devices.

2 Autism Spectrum Disorder Screening and Rehabilitation Device Categorization

We use the following mechatronics device categorization themes throughout this chapter:

1. Screening vs. therapeutic: screening devices are normally used as decision support or early warning systems. These devices collect data from subjects and data classification methods are later employed to distinguish between children with ASD and typically developing children (see [9] for further explanations and examples). Due to a lack of sufficient data, possible unpredicted circumstances, and heterogeneity among users, screening devices may not be used without skepticism about their accuracies. Rather, they would be used to warn families or clinicians about the need to perform further evaluation by experts.

    On the other hand, therapeutic devices are devices used in the disorder treatment process. They are mainly used as tools to assist therapists in their therapy sessions, such as RoboParrot [10]. On the other hand, there are new efforts to help parents in the therapy of their child by developing new devices easily usable by parents, such as the Joint Attention Bulbs (JAB) [11].
2. Wearable vs. non-wearable: by definition, wearable devices are connected to users to collect their data (the E4 wristband [12], for example). In contrast, the non-wearable devices are not worn by users, and users can collect data while away from them, such as the Microsoft Kinect [13]. It should be noted that the non-wearable devices are preferred to wearable ones, especially in children, since connecting devices to them is not easy. In kids with ASD, this would be more difficult since they may not cooperate due to their sensory oversensitivity.

3. Behavioral vs. biological/physiological data: subjects' data collection may be used as an invaluable tool for screening, evaluation, and also therapy adjustment. The biological/physiological data may include features such as heart rate, Blood Volume Pulse (BVP), and heart rate variability. On the other hand, the behavioral data refers to features such as gait patterns, pattern of engaging with an object such as a toy, or hand/body acceleration during activity or sleep. It should be noted that both behavioral and biological data can be collected via wearable or non-wearable devices. However, typically, behavioral data is mainly collected using non-wearable devices, while biological data is collected mainly using wearable devices.

Based on the applications of devices on ASD population, we categorized devices into *screening* and *therapeutic* classes. Furthermore, under each of the above categories, there are wearable/non-wearable and behavioral/biological subcategories.

3   Screening Devices for ASD

As mentioned earlier, children with ASD have specific behaviors or biological features which can be considered for the initial screening process. For instance, their repetitive/stereotypical behaviors, such as continuous tiptoe walking or turning [1], can be used to distinguish them from typically developing (TD) children. Consequently, mechatronic devices have been designed to collect either behavioral and/or biological data.

3.1   Screening Devices Designed to Collect Behavioral Data

As mentioned earlier, both wearable and non-wearable devices can be used to collect behavioral data. However, for collecting behavioral data, non-wearable devices are used more frequently than the wearable ones.

3.1.1   Non-wearable Devices

- Toy-based devices aim to capture the usage pattern of toys to distinguish between children with ASD and typically developing (TD) children. For example, the Intelligent Toy Car (ITC) [9] is equipped with a 3-axis accelerometer and collects the acceleration information while it is used by a child. Afterwards, five features are extracted from the collected data which are used to distinguish children with ASD from TD ones. The first version of ITC used a Nintendo Wiimote to collect acceleration data and transfer it to a remote computer for processing (Fig. 13.1). The new version of the car uses an accelerometer on a raspberry Pi board to

Fig. 13.1 The intelligent toy car equipped with a Nintendo Wiimote to collect acceleration information for screening. The car is equipped with flashing LEDs to make it more attractive for children, especially the children with ASD

collect and process data. The current accuracy in differentiating children with ASD from TD children is over 83 % on a group of 25 children with ASD and 25 TD children [9]. Interestingly, gender of the child has not been found to be a moderator or mediator, and thus the system can be used for screening both genders. The new version of the car also includes shaft encoders to collect the rotation of the car's wheels to include them into the classification process. It is believed that the accuracy can be improved by collecting more data and by including the shaft encoder data in the classification process.

- The gait pattern analyzer devices are designed to recognize the children's gait from which the difference between TD children and children with ASD can be determined. For instance, the UT gait analyzer system [13] uses an RGB-D camera such as Microsoft Kinect to detect a subject's body skeleton (Fig. 13.2) and keep track of the skeleton's coordinates over time. The subject's gait is then recognized from this data, and the difference between tiptoe walking and normal walking is determined. This allows the system to determine if a subject performs a typical walking motion or not. This system can also be used to determine other stereotypical motions such as flapping, which is typical in a group of children with ASD. It should also be noted that an advantage of this system is that it is markerless, i.e., there is no need to put markers on the body of subject to be able to determine his/her gait.
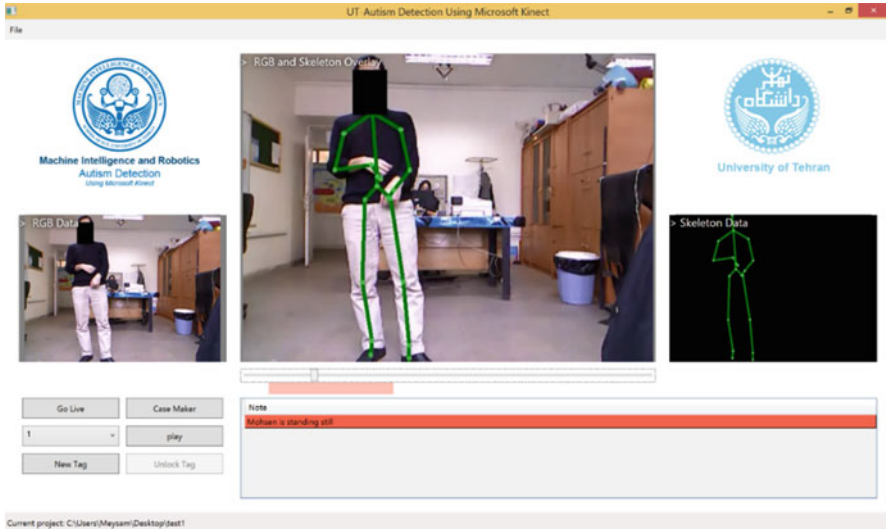
Fig. 13.2 The UT Gait Analyzer's user interface. The *left window* shows the actual camera view, the *middle window* shows the camera view with the detected skeleton superimposed on it, and the *right window* shows the skeleton by itself. The face of the subject is covered automatically to protect the privacy of subjects. The system allows a user to tag the video frame by frame or cut a specific part of it not needed

- Robot-based screening relies on a robot interaction with subjects while collecting behavioral data. The data is used to distinguish between TD children and ASD children based on differences between their behavioral features. For instance, RoboParrot [14] is a parrot-like robot which is used to interact with children and adults. The robot is capable of performing small movements, recognizing faces and colors, and features voice interaction. The robot was used in interactions with TD children and children with ASD, and showed that the ASD population interacts differently with the robot compared to the TD children. The differences in these interactions can be captured using a camera and processed by vision-based algorithms. QueBall [15], which is a ball shape robot, is another robot-based screening device designed to interact with children, and the interaction can be analyzed to determine whether or not the child has ASD or not.
- Eye trackers are devices developed to capture the gaze movement of subjects. The non-wearable version is designed to be placed in front of subjects, on a dedicated frame or attached to a target monitor. For instance, an eye tracker has been used to assess the relationship between the severity of ASD with the average fixation duration when seeing a face image [16]. It should be noted that the use of non-wearable eye trackers is limited since subjects' faces should be in the specific field of view of the tracker to be able to perform tracking.

The above three examples fall into non-wearable devices. In the following, a few examples of wearable ones are discussed briefly.

3.1.2 Wearable Devices

- Wristband devices are designed to be worn similar to wrist watches. These devices are designed to collect both behavioral and biological data, such as acceleration and heart rate, respectively. For instance, the E4 wristband [12] includes a 3-axis accelerometer to capture motion-based activities. This data may be used to determine the differences in the motion activities of children with ASD with respect to normal children. Another example is wireless accelerometers attached to children's bodies to recognize stereotypical motor movements such as hand flapping and body rocking [17].
- The wearable eye trackers are normally placed on an eyeglass frame and worn by subjects as normal eyeglasses. Campolo et al. [18] performed an experiment in which an eye tracker was place on a sport cap such that children can wear it easily. One of the uses of eye trackers in ASD screening is the joint attention evaluation. Joint attention refers to the shared focus of two individuals on an object, which is a fundamental capability in the learning process of children. Unfortunately, many children with ASD show deficits in this capability. In a case of joint attention evaluation [19], the subjects' joint attention is evaluated while eye gaze of the examiner is tracked by a wearable eye tracker and at the same time a camera is used to track the gaze of the child under examination, and their joint gaze is measured to evaluate the subject's joint attention.

3.2 Screening Devices Designed to Collect Biological Data

The following devices are used to collect biological/physiological data. These devices are mainly in the wearable category due to the fact that the biological/-physiological data may not be collectable without being connected to subjects.

- Wrist band devices tend to collect biological/physiological data. For instance, the E4 wrist band [12] collects heart rate and heart rate variability based on the blood volume pulse measured using photoplethymography sensors. In addition, it collects arousal and other features related to stress, engagement, and excitement through a electrodermal activity sensor. Furthermore, skin temperature is calculated using an infrared thermopile. The researchers hope that collecting this information will help them to screen children with ASD.
- Head band devices tend to collect brain-related activities. There are many technologies used for collecting brain activities, such as those collecting near-infrared waves [20], or electroencephalography (EEG) based devices [21] which also collect brain activity and provide additional information on the activation of different brain areas. The examination of the brain activities in a subject can show whether a specific brain region shows atypical activation, or if there are problems in the connectivity between different regions of the brain [21]. For example, in [22] it is shown that the brain connectivity and brain networks dynamics in children with ASD is different than the TD children.

- The Language ENvironment Analyzer (LENA) system uses signal processing techniques to analyze the vocalization of children. The empirical data [23, 24] have shown that there are rich vocal discriminations between normal and children with ASD, and that [24] specific features in the voice of children with ASD may discriminate them from TD children. In addition, researchers tend to implement similar algorithms on handheld devices such as smart phones or dedicated devices. The wide spread use of smart phones, their sensor collection, and their connectivity to the Internet make them a very suitable device for collecting and processing data.

4 Therapeutic Devices

ASD rehabilitation process is a very complex task and needs tremendous efforts from therapists and parents. Thus, introducing complementary devices to help and ease the process are necessary and welcomed by caregivers. There have been many attempts to design software, robots, and devices to help in the rehabilitation process. Here, we introduce a few examples that we categorize into robots and devices. Since there may not be a clear distinction defined between devices and robots, these terms are used based on how the inventors named their inventions.

4.1 Robot-Based Therapeutic Devices

In this approach, robots are used to help in therapy process. In general, due to the attractiveness of using robots [25], children may show interest in engaging with robots, making the rehabilitation process easier. In general, three categories of robots are used: (a) humanoid robots, (b) animal-like robots, (c) robotic devices.

Nao (Fig. 13.3), as a humanoid robot, has been widely used in ASD rehabilitation due to its available features and price. It has largely been used for imitation learning. Consequently, it has been employed conjunctly with devices such as Microsoft Kinect to detect a child's motion pattern and try to mimic it or vice versa [26, 27].

Dickstein-Fischer et al. [28] designed and developed a robot capable of screening and rehabilitating of children with ASD. The robot has two eyes with 3 degrees of freedom (DOF), with 1-DOF eyelids, a 3-DOF head, a 1-DOF beak, and 1-DOF wings alongside the ability of recognizing head orientation and eye-gaze tracking suitable for interacting with children. These capabilities would allow the robot to be used by a remote caregiver to screen and perform a therapy process.

Since children with ASD often favor animals, therapy using animals may be a promising approach [29]. Designing and using animal-like robots has therefore been an approach in the rehabilitation process [30]. For example, RoboParrot [10] is used in a turn-taking therapy scenario in which the therapist uses the robot as the second person under therapy. The subject and the robot are supposed to follow the rules and
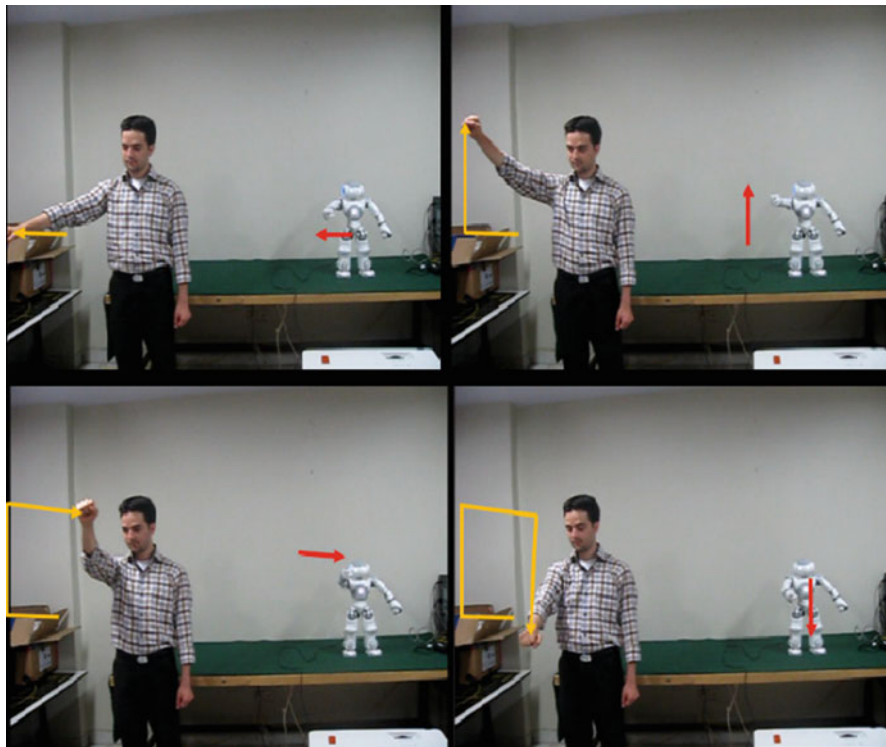
Fig. 13.3 Nao, on the right, imitating a human [23]

only answer when it is their turn. It can be imagined that this eases up the therapy process since there are many cases in which there is no other person to take part in this process, thus the robot steps in. Furthermore, the study shows that children respond well to the robot, which facilitates the therapy session.

Keepon is another animal-like robot that has been used [31] to activate the intact motivation of children with ASD. It is shown [31] that children started sharing their interests and feelings with others during and after using Keepon. Although Keepon is a very simple robot capable of showing its gaze direction and basic emotions, it was capable of easing up the therapy process.

Pleo is an off-the-shelf toy with 13 prerecorded scripts and 10 preprogrammed social behaviors. It has been used on several children with ASD and has been shown that the children spoke more while in interaction with Pleo than with adults [32].

Another example is a cat-like robot [33] which provides several needed features for ASD therapy such as eye contact, hugging, body movement, and reaction to negative behaviors. The furry coat of the robot can help to attract children. Furthermore, the facial expressions and the body movements are features implemented in the robot that can be used in therapy scenarios.
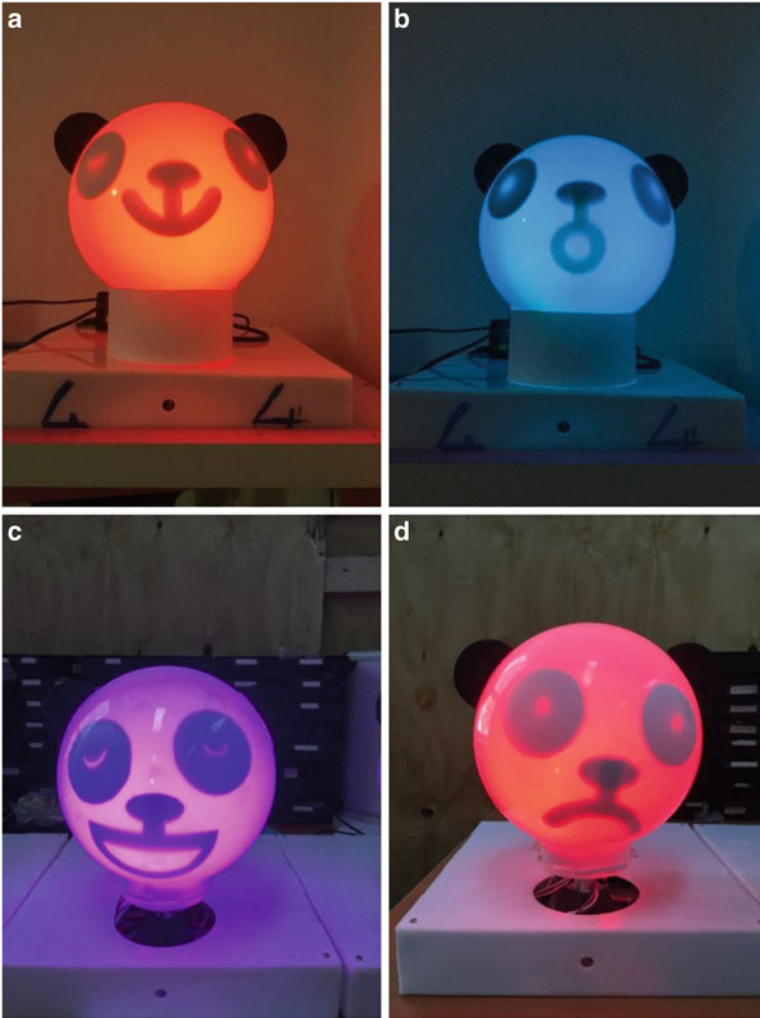
Fig. 13.4 The joint attention bulbs with different colors and emotions. The user can turn on and off these features such as sad or surprised faces (**b**, **d**). These appear when a child has not looked at the bulbs, while the happy faces (**a**, **c**) appear once the child has looked at the bulbs

4.2 Dedicated Therapeutic Devices

In this approach, a device is designed to take part in therapy scenario. For instance, joint attention bulbs [11] are designed to be used in joint attention therapy by therapists or by parents (Fig. 13.4). The bulbs are controlled through an Android application. Thus, therapists or parents can use them to teach children to look at what they are pointing to. In a typical therapy scenario, three or four bulbs are located

around the therapy room, and the therapist would point toward one of the bulbs. If the child under therapy does not follow the pointing direction, the therapist can stimulate him/her by switching on the desired bulb. If this is not enough to attract the child, the therapist can activate the bulb's designed voice system to attract the child. If the child still does not pay attention, the combination of the light and sound is used to attract the child toward the bulb that is pointed by the therapist. Using this approach, the child is stimulated through multiple media, which is expected to yield greater impact than with the traditional methods. Initial results, through a single subject experiment design on two children with ASD, show significant improvement in their joint attention [11].

Another example is a wearable EMG device designed to recognize facial expressions [34]. This helps automatic emotion detection in the process of teaching emotion understanding. The automatic emotion detection allows robots to detect a child's emotion and respond accordingly in the emotion understanding therapy scenarios.

Microsoft Kinect is an off-the-shelf motion capture device and has been widely used in ASD intervention process [26, 27]. Such motion capture devices are either used with avatars implemented in computer games and programs [26], or with robots in therapy sessions [26, 27]. Since children with ASD are often fascinated by technology, they tend to interact with such devices easily. One of the interesting findings presented by Meghdari's team [27] is that such interactive games, using motion capture devices, can provide a medium for child–parent play which can improve the relation between children with ASD and their parents.

LENA system [35] has also been used to count the number of words used by children with ASD with or without therapy to show the importance of vocal therapy on these children. This use of LENA system shows the possible application of such devices for evaluation and assessment.

The standard approach in ASD rehabilitation assessment is to observe children and rank them in all the relevant areas of impairment, which is time consuming and very subjective. Furthermore, the expert-based evaluation is usually performed every 6 months or so. In contrast, the use of such devices would allow regular assessment at lower price and with easier implementation.

5   Conclusion

In this chapter, a set of mechatronic devices for screening and rehabilitation of children with ASD is discussed. The prevalence of ASD and its difficulties in screening and therapy processes necessitates development of new mechatronic devices which are accurate and robust enough to be used by caregivers and the general public. The design of screening methods and devices should be focused on early age stages of children's life (under 2 years of age). Initial attempts at developing such devices and methods have begun, such as the voice-based ASD screening [22]. An important characteristic of mechatronic devices, designed to

serve as an ASD screening system, can be their use in children's natural setup by parents. This would extend their use beyond medical centers and make these devices usable all around the world. This is of key importance since there are many countries that do not have widespread medical service centers and experts to diagnose ASD as early as possible.

Additionally, the important phase after the diagnosis step is therapy, which is very costly and requires high commitment and engagement from families. Furthermore, extensive therapy is needed to increase the efficacy of treatment while intervention support devices can be helpful to therapists in easing up the long therapy process. On the other hand, these systems may enable parents to perform therapy at home on their own, as well maintain the ability to reduce the cost and overhead of ASD therapy significantly. The inclusion of smart phones and related applications, which benefit from the sensors and actuators embedded in these devices, can further reduce the cost and increase the friendliness of the systems for parents and caregivers.

References

1. American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Association, Arlington
2. Szatmari P (2003) The causes of autism spectrum disorders. Multiple factors have been identified but a unifying cascade of events is still elusive. Br Med J 326(7382):173–174
3. Golestan S, Soleiman P, Taban R, Moradi H (2016) A comprehensive review of technologies used in screening, rehabilitation, and assessment of autism spectrum disorder (ASD). MIR_TechReport 94-10-29/1, School of Electrical and Computer Engineering, University of Tehran
4. Goldstein S, DeVries M (2013) Autism spectrum disorder enters the age of multidisciplinary treatment. In: Gledstein S, Naglieri JA (eds) Interventions for autism spectrum disorders, translating science into practice. Springer, New York, pp 3–18
5. Samadi SA, McConkey R (2011) Autism in developing countries: lessons from Iran. Autism Res Treat 2011:145359
6. http://www.cdc.gov/ncbddd/autism/hcp-screening.html
7. Fletcher PC, Markoulakis R, Bryden PJ (2012) The cost of caring for a child with an autism spectrum disorder. Compr Pediatr Nurs 35:45–69
8. http://espressif.com/en/products/esp8266/
9. Ekhtiari S, Moradi H, Pouretemad HR, Araabi B (2012) Early autism screening using an intelligent toy. In: The 5th international congress of child and adolescent psychiatry
10. Soleiman P, Moradi H, Mahmoudi M, Pouretemad HR (2015) Using a parrot like robot for speech therapy of children with autism. In: The 4th basic and clinical neuroscience congress, Tehran, Iran, 23–25 Dec 2015
11. Pourmemar M, Moradi H, Pouretemad HR (2015) The design, development, and evaluation of a system for joint attention therapy for children with autism. In: The 4th basic and clinical neuroscience congress, Tehran, Iran, 23–25 Dec 2015

12. Hernandez J, McDuff DJ, Picard RW (2015) BioInsights: extracting personal data from "still" wearable motion sensors. In: Proceedings of body sensor networks conference, MIT, Cambridge, June 2015

13. Ebrahimi M, Feghi M, Moradi H, Mirian M, Pouretemad HR (2015) Distinguishing tip-toe walking from normal walking using skeleton data gathered by 3D sensors. In: The 3rd international conference on robotics and mechatronics, Tehran, Iran, pp 450–455

14. Soleiman P, Moradi H, Mahmoudi M, Pouretemad HR (2015) The design, development, and deployment of RoboParrot for screening autistic children. Int J Soc Robot 7:513–522

15. Golliot J, Raby-Nahas C, Merat Y, Beaudoin A, Côté B, Duclos C (2015) A tool to diagnose autism in children aged between two to five old: an exploratory study with the Robot QueBall. In: HRI'15 extended abstracts. Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction, pp 61–62

16. Amirhafteran S, Irani B, Pouretemad H, Samaknejad N, Moradi H (2015) The average fixation duration on the face is the indicator of communication disorder in autistic children. In: The 6th international conference of cognitive science, Tehran, Iran, 27–29 Apr 2015, p 79

17. Albinali F, Goodwin MS, Intille SS (2009) Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum. In: Proceedings of the international conference on ubiquitous computing, pp 71–80

18. Campolo D, Taffoni F, Shiavone G, Laschi C, Keller F, Guglielmelli E (2008) A novel technological approach towards the early diagnosis of neurodevelopmental disorders. In: The 30th annual international IEEE EMBS conference, Vancouver, Aug 2008

19. http://www.gregoryabowd.com/research/projects/3d-eye-gaze-tracking-estimate-joint-attention

20. Hemmati N, Setarehdan SK, Ahmadi Noubari H (2012) Multi-channel near-infrared spectroscopy (NIRS) system for noninvasive monitoring of brain activity. In: Proceedings of the IEEE-EMBS international conference on biomedical and health informatics (BHI), Hong Kong and Shenzen, China

21. Coben R, Mohammad-Rezazadeh I, Cannon RL (2014) Using quantitative and analytic EEG methods in the understanding of connectivity in autism spectrum disorders: a theory of mixed over- and under-connectivity. Front Hum Neurosci 8:45

22. Buard I, Rogers SJ, Hepburn S, Kronberg E, Rojas DC (2013) Altered oscillation patterns and connectivity during picture naming in autism. Front Hum Neurosci 7:742

23. Xu D, Gilkerson J, Richards J, Yapanel U, Gray S (2009) Child vocalization composition as discriminant information for automatic autism detection. In: Proceedings of the IEEE international conference on engineering in medicine and biology, Sep 2009

24. Ebrahimi Motlagh H, Moradi H, Pouretemad HR (2013) Using general sound descriptors for early autism detection. In: The 9th Asian control conference (ASCC), pp 1–5

25. Diehl JJ, Schmitt LM, Villano M, Crowell CR (2012) The clinical use of robots for individuals with autism spectrum disorders: a critical review. Res Autism Spectr Disord 6(1):249–262

26. Taheri AR, Alemi M, Meghdari A, Pouretemad HR, Holderread SL (2015) Clinical application of humanoid robots in playing imitation games for autistic children in Iran. Soc Behav Sci 176:898–906

27. Greczek J, Kaszubski E, Atrash A, Mataric M (2014) Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders. In: The 23rd IEEE international symposium on robot and human interactive communication, pp 561–566

28. Dickstein-Fischer L, Alexander E, Yan X (2011) Affordable compact humanoid robot for autism spectrum disorder in children. In: 33rd annual international conference of the IEEE EMBS, pp 5319–5322

29. Nimer J, Lundahl B (2007) Animal-assisted therapy: a meta-analysis. Interact People Anim 20(3):225–238

30. Stanton CM, Kahn PH, Severson RL, Ruckert JH, Gill BT (2008) Robotic animals might aid in the social development of children with autism. In: The 3rd ACM/IEEE international conference on human-robot interaction, pp 271–278

31. Kozima H, Nakagawa C, Yasuda Y (2007) Children-robot interaction: a pilot study in autism therapy. Int J Prog Brain Res 164:385–400
32. Kim ES, Berkovits LD, Bernier EP, Leyzberg D, Shic F, Paul R, Scessellati B (2013) Social robots as embedded reinforcers of social behavior in children with autism. J Autism Dev Disord 43(5):1038–1049
33. Lee BH, Jang J, Mun K, Kwon JY, Jung JS (2014) Development of therapeutic expression for a cat robot in the treatment of autism spectrum disorders. In: 11th international conference on informatics in control, automation and robotics (ICINCO), 2014, vol 2, pp 640–647
34. Hirokawa M, Funahashi A, Itoh Y, Suzuki K (2014) Design of affective robot-assisted activity for children with autism spectrum disorders. In: Robot and human interactive communication, 2014 RO-MAN: the 23rd IEEE international symposium
35. Gilkerson J, Richards JA. Impact of adult talk, conversational turns, and TV during the critical 0-4 years of child development. LENA foundation technical report ITR-01-2

Chapter 14
Electrochemically Derived Oxide Nanoform-Based Gas Sensor Devices: Challenges and Prospects with MEMS Integration

P. Bhattacharyya, K. Dutta, and P.P. Chattopadhyay

Abstract Since their invention, advents in structural evolution of semiconducting metal oxide-based chemical sensors relied much on the controllable surface to volume ratio of different nanoforms as the governing factor for sensing performance. Among the established synthesis techniques, electrochemical anodization has been found to be the promising one to develop the porous sensing layer with precise controllability and repeatability. The method is used either to create controlled porosity on an existing oxide surface or to grow porous oxide from corresponding metal. Recently, the technique has been successfully extended for development of sophisticated oxide structures like nanotubes, nanowires, and nanodots. The scope of the present overview concerns electrochemical anodization process, encompassing the effect of texture influencing factors, followed by the sensing performance of such films/nanoforms. Potentiality of different device structures (like resistive, Schottky, Metal–Insulator–Metal (MIM) and Metal–Insulator Semiconductor (MIS)) employing such electrochemically grown layer of oxide nanoforms, as the sensing element, have been elaborately explored. The sensor characteristics like response magnitude, response time and recovery time for the oxidizing and reducing chemical species have been critically discussed with particular emphasis on respective advantages and bottlenecks of the device structures. Finally, the chapter summarizes the salient features, prospects, and challenges of the electrochemically grown nanostructures for possible integration with MEMS technology.

P. Bhattacharyya (✉) • K. Dutta
Department of Electronics and Telecommunication Engineering, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, West Bengal 711103, India
e-mail: pb_etc_besu@yahoo.com; pb@telecom.iiests.ac.in

P.P. Chattopadhyay
Department of Metallurgy and Materials Engineering, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, West Bengal 711103, India

1   Introduction

Semiconductor oxide-based chemical sensors during their evolution from primitive "Taguchi" type sensor to sophisticated MEMS based ones, have embraced different nanoforms as the sensing layer owing to the unique features of such nanostructures. For a large variety of target species, the sensing performance is predominantly governed by "adsorption and desorption" kinetics on the oxide surface of these chemoresistive sensors [1–3]. Porosity of the sensing layers play a pivotal and decisive role for the efficient (detection capability to very low ppm level with high sensitivity) and fast detection of gaseous/vapor species. For compact films (having less porosity) the gas interaction phenomena is restricted only to a few upper atomic layers of the film and cannot penetrate into the deep, which eventually leads to lower electron contribution capacity (as a result of gas interaction process) and therefore towards lower sensitivity [3, 4]. On the other hand, for porous sensing layer, usually the sensitivity is higher as the gas interaction takes place both on the surface and within the pores. Moreover, it has recently been established that the porosity-originated nanoforms like nanotubes [5–19], nanowires [20, 21] and nanodots [22] are extremely efficient gas-sensing structures owing to their high surface to volume ratio. Figure 14.1 briefly describes the different nanoforms as the gas sensors.

Electrochemical anodization has already been established as an effective and efficient route for fabrication of ordered nanopores and nanoforms on semiconducting oxide surface [5–35]. The method, depending upon the type and the relative concentration of electrolyte(s), bias current/voltage, temperature of bath, anodization time etc., offers a large variety of structures with different dimensionality
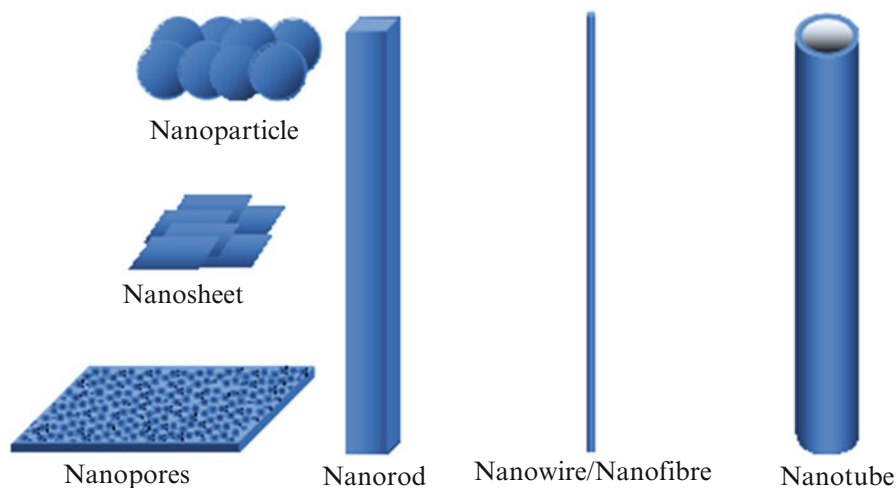
Fig. 14.1   Different types of nanoforms employed for gas sensing

and distribution having wide range of structural and electronic property variations. Usually, there are two methods for such pore formation. In the first approach, a semiconducting oxide grown/deposited by other means (non-electrochemical) is used as the starting material and electrochemical anodization is subsequently employed to create desired structure on that [24]. In the second approach, the native metal is used as anode in such a way that the structure grows on the metal itself (oxide formation combined with pore creation) [4, 22].

Different gas sensor device configuration, (as listed in Table 14.2) employing anodically grown semiconducting metal oxide sensing layer, have been critically reviewed. In the penultimate section, the prospects and challenges of integrating such nanoforms (electrochemically derived) with the MEMS technology is discussed with a comprehensive conclusion summarizing present status and future trends of the technology.

2   Electrochemical Anodization: Mechanism and Attributes

Anodization is a special case of electrochemical method to oxidize the metals. In this case, an oxide film can be grown on certain metal or etch the surface of certain oxide material employing an anodizing cell where the metal work-piece acts as the anode and an electrochemically inert conducting element (like Pt, graphite, etc.) as the cathode. The electrolyte(s) is chosen in such a manner that it reacts chemically with the anode under the influence of an applied bias. A reference electrode, usually saturated calomel (also inert to electrolyte), is used (optionally) to minimize the effect of the localized cells [4, 24, 27, 36–38].

For nanostructured sensing layer preparation, electrolyte is selected in a fashion that the grown/existing oxides can be dissolved in the electrolyte at a slower rate than the rate of oxide formation on anode [39]. The surface morphology/roughness, distribution, shape, and size of the nanoforms are highly dependent on different electrolytic medium as well as on the metal. Moreover, bias voltage, temperature, anodization time and the use of ultraviolet ray during anodization have a great impact on surface morphology of the resultant film [16, 18, 24].

2.1   Mechanism

Nanostructured (viz. nanoporous and nanotubes) array formation by the anodic oxidation technique is the result of three concomitant processes; (i) electrochemical oxidation, (ii) field-assisted dissolution and (iii) chemical dissolution [40–43]. Due to the application of a constant potential, the metal oxide layer starts to grow on the metallic surface and transforms slowly into a dense layer when the current through the electrolyte gets saturated [44]. Due to applied field between the anode

and cathode, metal ions (M^{2n+}) are diffused into the solution by the field driven ion diffusion process as shown by Eq. (14.1) [1].

$$M \rightarrow M^{2n+} + 2ne^- \tag{14.1}$$

$$2nH_2O \rightarrow 2nH^+ + 2nOH^- \tag{14.2}$$
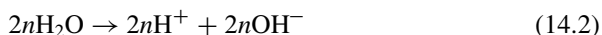
Oxide layer (MO_n) is deposited on the metal substrate by the chemical reaction between diffused M^{2n+} ion and OH^- or O^{2-} ions, according to the anode reactions shown in Eqs. (14.3) and (14.4) [44]. MO_n is also formed from $M(OH)_{2n}$ by the condensation reaction [Eq. (14.5)] [42].

Anode Reactions:

$$M^{2n+} + nO^{2-} \rightarrow MO_n \tag{14.3}$$

$$M^{2n+} + 2nOH^- \rightarrow M(OH)_{2n} \tag{14.4}$$

$$M(OH)_{2n} \rightarrow MO_n + nH_2O \tag{14.5}$$
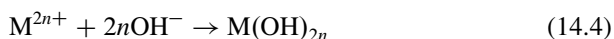
Hydrogen evaluation in the cathode surface is shown by Eq. (14.6).

Cathode Reaction:

$$2nH^+ + 2ne^- \rightarrow nH_2 \uparrow \tag{14.6}$$

Initially a very high current under the influence of large applied bias passes through the anode and subsequently oxide formation is initiated as shown in Fig. 14.2a. However, with the passage of time as the thickness and the compactness of the grown oxide layer increases, the resistivity of the anode increases thereby reducing
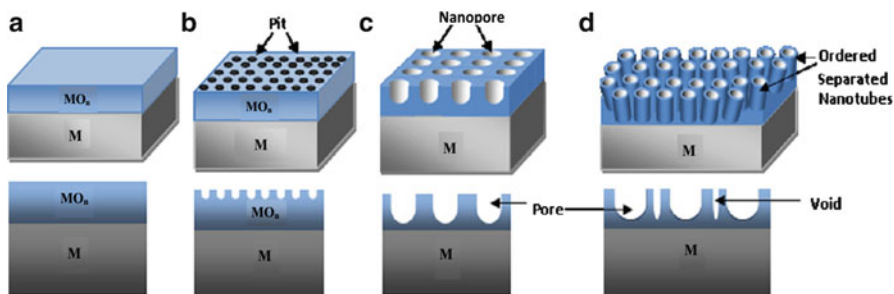
Fig. 14.2 Schematic of the nanopore and nanotube structure formation by the electrochemical anodization process (above: 3D view and below: corresponding 2D view), (**a**) oxide growth (**b**) pit formation (**c**) ordered nanopore formation (**d**) nanotube formation

the current through the cell which is manifested as a decreasing trend in the current with time as authenticated by different researchers [19, 40, 44]. Grown oxide layer on the metal surface is not uniform due to the presence of defects, local stress, etc. When the applied potential crosses the dielectric breakdown voltage of the grown oxide layer, sparking happens at the weak point (comparatively thin layer) of the film and the pores are created at the weak points of the oxide layer [40, 41, 45]. This type of pore formation is the result of field-assisted dissolution. Simultaneously, localized small pits are created due to the chemical dissolution process. This is a low-field phenomenon which is mostly responsible for nanostructure formation (Fig. 14.2b). Localized pits behave like the centers for pore creation and distributed almost homogeneously over the TiO_2 surface [44] as shown in Fig. 14.2b. These pits start to grow towards the bottom direction (towards metal substrate) and form the porous structure. Fluorine ions (F^-) have the ability of permanent chemical attacks (dissolution) over the grown metal oxides (like TiO_2, V_2O_5, ZrO_2, etc) and form highly soluble $\left[MF_6^{2-}\right]$ complex and also bear major responsibility for ordered pit formation on the oxide surface [42]. Beside the grown metal oxide (MO_n) layer, the metal hydrated layer ($M(OH)_{2n}$) and metal cations (M^{2n+}) are also affected directly by the F^- ions and forms soluble $\left[MF_6^{2-}\right]$ complex according to the following reactions [Eqs. (14.7)–(14.9)] [44, 46]. Due to this reason, fluoride-based electrolytes like NH_4F, HF, etc. are popular ones for ordered nanostructure formation. Porous structure formation followed by the nanotube array synthesis is depicted pictorially in Fig. 14.2c, d.

$$MO_n + mF^- + 2nH^- \rightarrow MF_m^{n-} + nH_2O \qquad (14.7)$$

$$M(OH)_{2n} + mF^- \rightarrow MF_m^{n-} + nOH^- \qquad (14.8)$$

$$M^{2n+} + mF^- \rightarrow MF_m^{n-} \qquad (14.9)$$
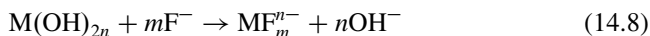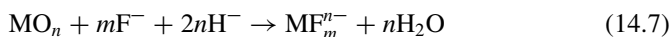
Fluorine-free electrolytes like $(NH_4)_2SO_4$ [39] and H_2SO_4 [24] are also well investigated for porous structure formation. Sulfate ions (SO_4^{2-}) attack the grown metal oxide and hydrated layer in the anodic surface and dissolve the oxide layer and create the pore on the oxide surface according to the Eqs. (14.10) and (14.11).

$$MO_n + mSO_4^- \rightarrow M(SO_4)_m + nO^{2-} \qquad (14.10)$$

$$M(OH)_{2n} + mSO_4^- \rightarrow M(SO_4)_m + nOH^- \qquad (14.11)$$
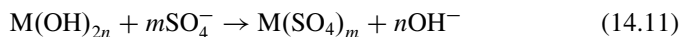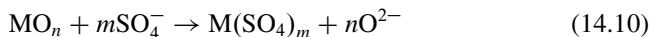
UV-assisted electrochemical anodization is another important technique for excess number of pore generation on the grown oxide surface. The electrochemical etching process upon UV illumination is due to the generation of excess holes and electrons in the valance band followed by the excitation of electrons to the conduction band of the active oxide insulator [24, 47]. These generated charge carriers

enhance the ionic conduction through the electrolyte and increase the oxide etching rate under a constant potentiostatic bias, thereby producing the porous structure. Reaction between Metal oxide (or hydrated layer) and sulfate ions $[(SO_4)^{2-}]$ by the generation of excess holes are shown in Eqs. (14.12) and (14.13).

$$MO_n + mSO_4^- + 2nh^+ \rightarrow M(SO_4)_m + nO_2 \qquad (14.12)$$

$$MO_n + mSO_4^- + 4h^+ \rightarrow M(SO_4)_m + nH_2O + \frac{n}{2}O_2 \qquad (14.13)$$
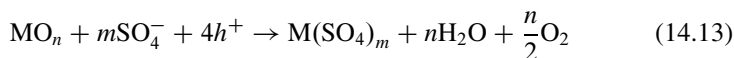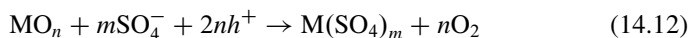
2.2 Effect of Anodization Parameters on Surface Morphology

In the field of chemical sensor, anodically grown metal oxides, like TiO_2, ZnO, WO_3, SnO_2, Al_2O_3 and Nb_2O_5, have so far been investigated to sense different target species like H_2, NO_2, CO, CH_4, HCHO, CH_3COCH_3, NH_3, H_2S, C_2H_5OH and CCl_4 [3–5, 7, 8, 13, 15, 17, 35, 48]. In all such reports, variation of the anodization parameters, viz. electrolyte (concentrations and compositions), time, applied bias, bath temperature etc. and their effects on the morphological and the structural aspects have been observed. The classification of different types of nanoforms, synthesized through the single step anodization is shown in Fig. 14.3. Detailed discussion of such governing factors on oxide morphology is presented in the following subsections. However, it is revealed that, most of the reports are focused on TiO_2 [5–19, 23–26, 33] and ZnO [20, 27, 36–38].
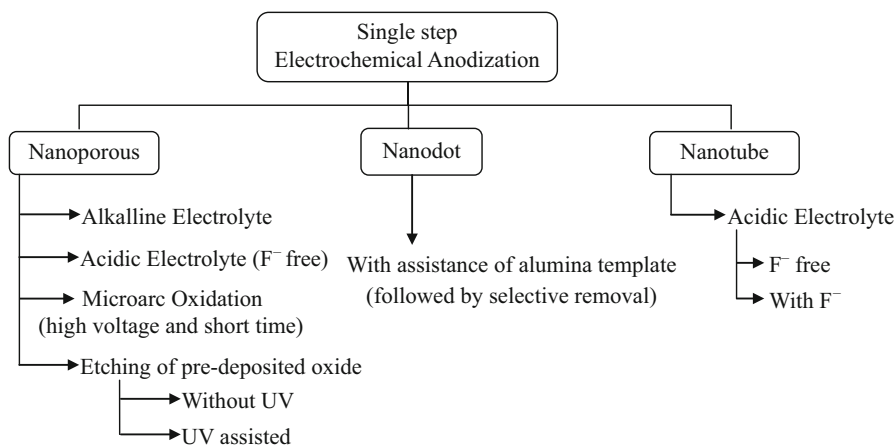
Fig. 14.3 Possible nanoforms via single step anodization with key process variations

2.2.1 Anode Material

In most of the cases, anodization process is performed using metallic substrate directly, as anode, to synthesize the respective oxide of that metal [5–21, 23, 28]. In indirect approach, a metal oxide layer is deposited on the non-metallic substrate (Si wafer, ITO, FTO, etc.) using different methods, prior to electrochemical etching procedure [22, 24]. Some of the authors reported metal bi-layer (double metal layer) as the anode (sputtered on Si wafer) to facilitate specific nanoform development or doping [22]. Types of oxide nanoforms are highly dependent on the starting material. It is observed that, most of the transition metal (Zn, Al, Ti, Nb, etc.) anodically oxidized into porous oxide layer [24–31, 33, 36–38]; whereas, nanotubular structure is particularly evident for titanium, hafnium and zirconium [5–19, 23, 49–51]. Titania nanotube is well investigated nanoform due to its chemical stability whereas zirconia nanotube is yet to be established as a reliable gas sensing element. Rani et al. [28] reported that, the pore dimension is larger in case of Nb_2O_5 than that of TiO_2, other parameters being same. Apart from these two nanoforms, equally spaced nanodots were resulted from anodization of metal bi-layer (with constant voltage anodization followed by second step anodization with voltage sweep to dissolve the oxide of top layer) [22]. It is found that, variation of current density determines the porosity as well as the pore diameter; while anodization time is responsible for thickness variation of the grown oxide layer [30]. Figure 14.4 depicts different nanostructures grown via electrochemical anodization incorporating different anode material.
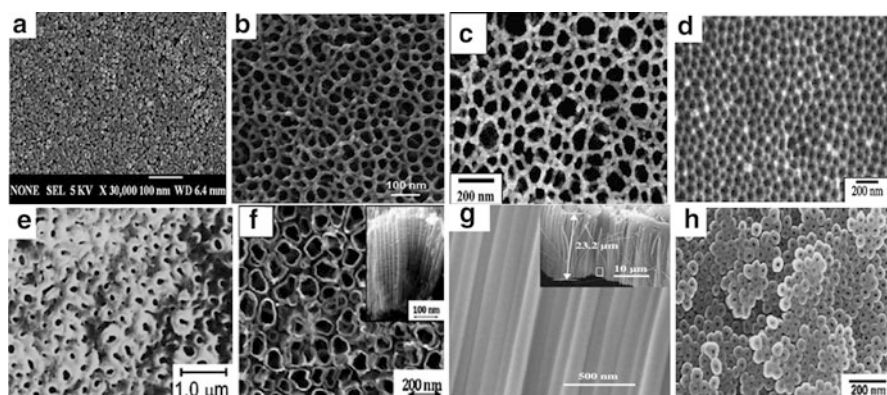
Fig. 14.4 Different anodic oxide nanoforms: (**a**) nanoporous ZnO, (**b**) nanoporous Nb_2O_5, (**c**) nanoporous SnO_2, (**d**) nanoporous Al_2O_3, (**e**) nanoporous TiO_2, (**f**) TiO_2 nanotube, (**g**) ZrO_2 nanotube (showing thickness of the film in *inset*), and (**h**) HfO_2 nanotube (reprinted with permission from [3, 19, 25, 27, 28, 35, 50, 51])

2.2.2 Electrolyte

Electrolyte (compositions and concentrations) used in electrochemical anodization determines the resultant morphology as well as properties of the anodically grown metal oxides [16, 19]. To synthesise metal oxides via electrochemical method, alkaline and acidic, both types of electrolytes are used. The effect of such electrolyte(s) on the morphology of the resultant film is discussed in the following subsections.

Acidic/Alkaline Electrolyte

Literatures review established that, alkaline electrolyte can only be used to synthesize nanoporous oxide layer [25, 26]; whereas acidic ones enable nanopore as well as nanotube formation depending on the process parameters (like anode material, applied voltage, anions present in electrolyte etc.) [5–29, 33, 34]. Hyodo et al. [29], experimentally showed the effect of electrolyte (acidic/alkaline) on variation of the pore dimension and the pore distribution of the grown oxide layer. It is also noticed that, alkaline (0.5 M NaOH) medium results in pores with poor ordering compared to its acidic (0.5 M H_2SO_4) counterpart [25, 29]. Moreover, average distance between the two pores is larger in case of alkaline electrolyte; i.e., relatively less porous oxide film with reduced surface roughness is evident for alkaline electrolyte [29]. The dimension and the ordering of pores are crucial for tuning gas sensing performance. Hence the use of acidic electrolyte is more popular than alkaline ones for gas sensor applications [4, 15, 16, 19, 27, 28, 31–34, 36, 38].

Weak and Strong Electrolytes

Depending on acidity, acidic electrolytes are further subdivided into strong (viz. H_2SO_4, HF, NH_4F) and weak (viz. H_3PO_4) electrolytes. Strong electrolytes are responsible for lower dimension of pores in large numbers whereas weak electrolytes result in fewer number of pores with larger diameters [29]. Besides, the use of fluorine containing strong electrolytes substantially promotes the formation of nanotubular structure [5–19]. In case of nanotube growth, mixing of the weak electrolyte (H_3PO_4) with the strong ones (HF/NH_4F) results in better wall separation (improvement in free-standing nature of the nanotube) between two nanotubes than those produced with only strong electrolyte (results in wall connected nanotubes) [52]. Hazra et al. [40] proposed the formation mechanism for such wall separated nanotubes, under the influence of mixed electrolyte (weak + strong one), considering continuous repulsion effect of cations (negatively charged) and their radial movement towards the electrolyte/oxide interface region. In addition, some reports are also focused on synthesis of nanotubular structure by anodic oxidation by using fluoride-free electrolyte (with Cl^- ion), though the application of these materials as gas sensor is yet to be reported [53, 54].

Aqueous/Non-aqueous Electrolytes

In general water is used as the solvent or as the medium of the electrolyte. This is also the main source of the oxygen required for oxidation of the respective anode material in electrochemical process [3, 4, 10, 11, 13, 21, 22, 24–27, 29–32, 34–36, 38, 48, 49, 55]. Besides, non-aqueous organic solvents like, ethylene glycol, glycerol, ethanol etc. are also used for this purpose [5–9, 12, 15–20, 28, 33, 50, 51]. Use of such medium improves control over the oxygen supply to anode during anodization and hence, variations in the amount of water in the organic electrolyte (ethylene glycol, glycerin, etc.) produce variations in the amount of oxygen vacancies, which is one of the key determining factors for tuning gas sensing performance [15, 16, 19]. It is also observed that, uneven outer wall (presence of annular ring) of nanotube is evident when aqueous electrolyte is used and in case of non-aqueous (mainly organic) electrolyte relatively smooth walls can be achieved [15, 52].

2.2.3 Bath Temperature

According to "Stoke-Einstein theory," diffusion coefficient (D) of a sphere like particle (of diameter a) depends linearly on ambient temperature (T in Kelvin) and viscosity (η) of the medium [56];

$$D = \frac{kT}{6\pi \eta a} \tag{14.14}$$

Increase in bath temperature, increases the ionic activity within the electrolyte. As a result, the oxidation process as well as the dissolution process increases. Combined effect of these two processes, varies the distribution, the size and the density of pores [26, 31]. To the contrary, nanotube wall thickness is decreased for increase in bath temperature [16]. Variation of anodization parameters and their effect on TiO$_2$ nanotube morphology was extensively studied by Hazra et al. [16] and the corresponding FESEM images are shown in Fig. 14.5.

2.2.4 Electrode Potential

Anodization is usually carried out in potentiostatic mode in most of the cases [5–24, 27, 28, 33, 34]. However, anodic oxidation applying current density was also investigated by some researchers [25, 26, 29–32]. Variation of potential drop between the anode and the counter electrode ("cathode" in case of two-electrode cell and "reference electrode" in three-electrode cell) results in variation of nanotube diameter in a linear fashion [10]. Jun et al. [48] combined high voltage for short duration (termed as microarc oxidation or MAO) to prepare porous titania film which exhibited different orientations of pore distributions.
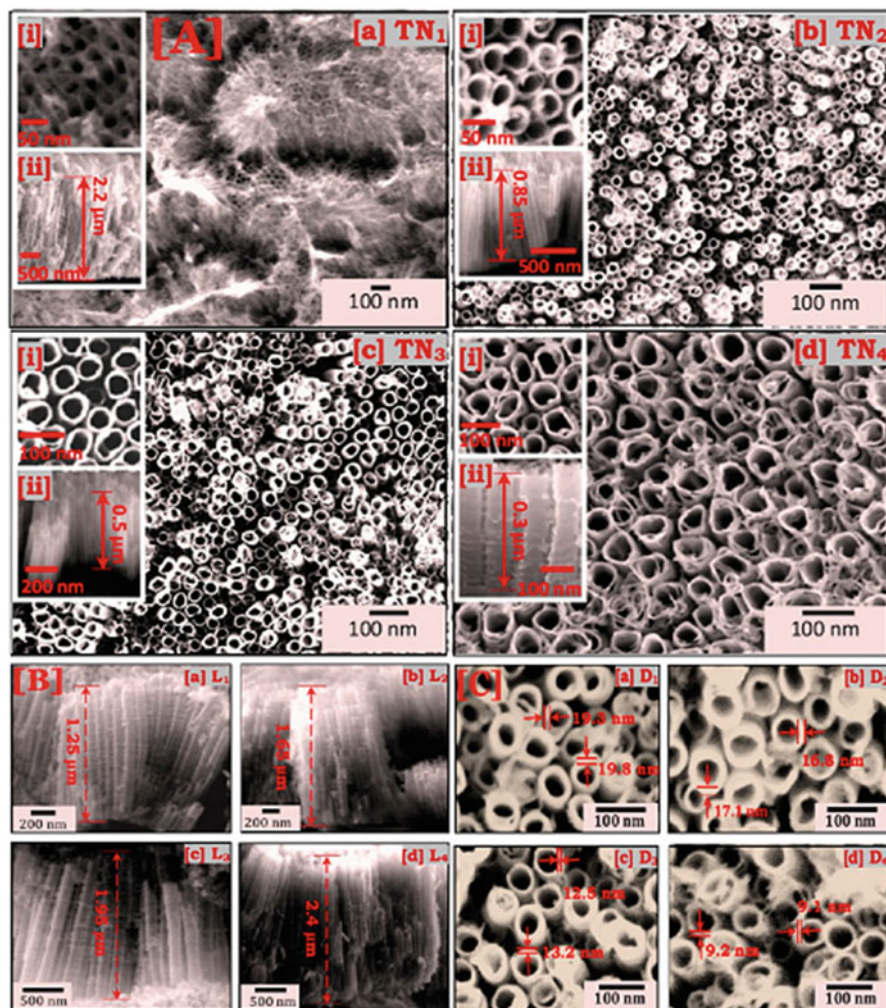
Fig. 14.5 Morphological variation of TiO$_2$ nanotube depending upon different anodization parameters: [A] Water content: (**a**) 0, (**b**) 2, (**c**) 5, and (**d**) 100 vol.% water in ethylene glycol-based 0.5 wt% NH$_4$F solution (room temperature for 2 h); [B] Anodization duration: (**a**) 4, (**b**) 8, (**c**) 12, and (**d**) 16 h (2 vol.% water in ethylene glycol-based 0.5 wt% NH$_4$F solution); [C] Bath temperature: (**a**) 42, (**b**) 57, (**c**) 72 and (**d**) 87 °C (2 vol.% water in ethylene glycol-based 0.5 wt% NH$_4$F solution) (reprinted with permission from [16])

2.2.5 Anodization Time

Increased anodization time increases the oxidation process towards oxide–metal interface which eventually increases the thickness of the grown oxide layer [16, 18]. Bhowmik et al. [18] illustrated that the length of the nanotube increased from

∼208 nm to ∼2.4 µm when the anodization time increased from 1 to 16 h. It is also observed that, the effective surface area increased as a function of tube length; but defect states got reduced for longer nanotube array [18]. The study shows that, the duration of electrochemical process increases effective surface area, which is desired for better sensing performance. However, it reduces the defect states present in the oxide, which deteriorates the sensing ability [18]. Hence, judicious optimization of anodization time is required for efficient tradeoff between the above competing factors.

2.2.6   UV Radiation

Application of ultra violet (UV) ray during anodization, increases the etching capability of the electrolyte which results in thinner oxide layer with larger pore diameter [24]. Further, in case of electrochemical etching of the pre-deposited oxide, enhanced vertical axis etch (without branching) was achieved with UV radiation which eventually manifests itself as nanorod-like structure [24].

Thus, by controlling different anodization parameters, the morphological features of the grown oxide layer can be varied easily. Table 14.1, comprehensively depicts the scenario of anodic growth/etching of metal oxides for potential application as gas sensor.

3   Sensing Performance of Different Nanoforms (Electronic Interaction Between Sensor and Target Species)

For chemical sensors, the surface morphology, surface to volume ratio to be specific, plays the most critical role in determining sensing performance [3, 15, 30]. Besides, the amount and the nature of defects determine the reactivity of the sensing surface with the target species [15, 16]. Hence, better sensing performance is achievable with judicious optimization among the amount of defects and the surface area. Also doping, by the catalytic metals, influences the sensor performance significantly [27].

The amount and time requirement for change of conductivity of the sensing surface due to chemisorptions of target species is represented as sensitivity and response time, respectively [1]. To improve these parameters, researchers improvised different device structures.

Planar/resistive type device (Fig. 14.6a) is the mostly used and the simplest device structure where two electrodes are either deposited on the top or at the bottom of the sensing layer [1, 19, 35]. For obvious reason, the distance between the electrodes determines the electron transportation path length and hence response time. In general, the resistance/current between these two electrodes is measured in presence and in absence of target species and the difference in resistance/current is quantitatively represented as the response magnitude.

Table 14.1 Different anodization conditions resulting in different nanoforms of various metal oxides

Structural morphology	Anodization specification					
	Anode material	Electrolyte	Electrical input	Duration	Bath temp (°C)	Reference
Microporous	Nb	Aq. H_3PO_4	100 mA/cm²	30 min	20	[29]
Sintered particles	Nb	Aq. NaOH				[29]
	Ti	Aq. NaOH				[25, 26]
Non-ordered nanoporous	Nb	Aq. H_2SO_4				[29]
	Ti	Aq. H_2SO_4			5, 20 and 30	[25, 26, 32]
		Aq. H_2SO_4	200 V	10 min	–	[31]
		Aq. H_2SO_4	20 and 50 mA/cm²	30 min	–	[48]
		Aq. HF	100 mA/cm²	30 min	20	[30]
						[32]
	Zn	Aq. Oxalic acid	10 V	–	RT	[4, 27, 36, 38]
Non-ordered nanocrystalline	Ti	Aq. H_2SO_4	10 V	30 min	RT	[24]
Ordered nanoporous	Sn/SiO_2/Si	Aq. Oxalic acid	6 V	15–90 s	5–25	[3]
	Ti	EG. NH_4F	5 V	60 min	RT	[33]
	Nb	EG. NH_4F	10 V	30/60 min	50	[28]
	W	Aq. NaF	20 V	10, 12.5, 15 min	–	[34]
Highly ordered nanoporous	Adhesive backed Al Tape	Aq. H_2SO_4	15 V	120 min	–	[35]
Nanoporous and nanowire skeleton	W	Aq. NaF	60 V	30 min	2	[55]

Nanoform	Material	Electrolyte	Voltage	Time	Temp	Ref
Coral-like porous		Aq. NaF& HF	40 V	60 s	–	[21]
Ordered nanodots	W/Ti/Si	Aq. $H_2C_2O_4$ and NH_4F	100 V	–	–	[22]
Nanowires	Zn	E. Oxalic acid	15–80 V	–	RT	[20]
Nanotube array	Ti	Aq. HF	12–20 V	25 min	–	[10]
		Aq. HF	20 V	45 min	20	[11]
		Aq. HF	3–20 V	–	0–15	[13]
		EG. NH_4F	45 V	30 min	25	[5]
		EG. NH_4F	60 V	60 min	–	[6]
				180 min	RT	[7]
		EG.NH_4F+ H_2O	20 V	120 min	RT	[15, 16, 19]
			20 V	4–16 h	RT	[16, 18]
			20 V	120 min	42–87	[16]
		G. NH_4F	OC to 30 V	3 h	–	[9]
		G.NH_4F	20, 30, 40, 50, 60 and 100 V	3 h	–	[12]
	Zr	EG. NH_4F+ H_2O	30 V	5 min	20	[51]
	Hf	EG. NH_4F+ H_2O	20 V	20 min	20	[50]
Nanoporous and nanotube		Aq. H_2SO_4+ NaF	10–60 V	60 min	RT	[49]
Pulverized nanotube	Ti	EG. NH_4F	30 V	Up to $I_A = 0$ A	RT	[8]

Aq. aqueous, *EG.* ethylene glycol, *G.* glycerol, *E.* ethanol, *RT* room temperature, I_A anodization current, *OC* open circuit
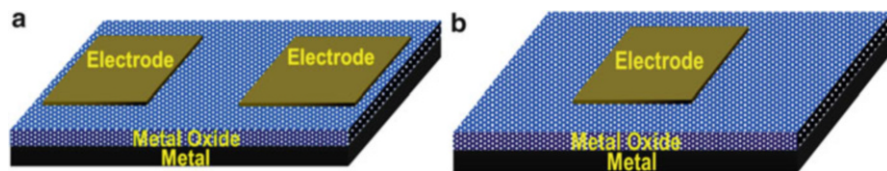
Fig. 14.6 Schematic of (**a**) resistive/planar and (**b**) metal insulator metal (MIM) gas sensor device configuration

On the other hand, in Metal-Insulator-Metal (MIM) configuration (Fig. 14.6b), the sensing layer is sandwiched between the two electrodes [18]. Generally, oxide forming metal of the sensing layer is used as one of the electrodes (usually Ohmic and bottom one). Thus in such cases, the electron transportation path is approximately equal to the thickness of the grown oxide which is eventually less than that in case of the planar device. From device structure perspective, Metal-Insulator (active)-Metal or MIM structure is more efficient for gas sensor application, as it offers much faster response time compared to the planar ones. It was supported experimentally by several researchers [24, 27].

Micro Electro Mechanical System or MEMS, based on silicon technology, finds suitability in the field of gas sensor as the silicon substrate serves as common platform for the sensor device as well as the electronic circuitry (required for processing the sensor signal) [57]. Also, facility of batch fabrication of such system lowers the cost substantially. Moreover, due to such miniaturized structure (reduced thermal mass of the substrate) and excellent thermal insulation between the heated active area and the substrate results in lower power consumption of MEMS-based integrated devices compared to the bulk substrate-based discrete ones. Integration of such MEMS technology with anodically grown oxide as the sensing layer paves the path of low power sensors with high packing density and sufficiently high sensitivity and fast response. In the Fig. 14.7, an illustration of MEMS-based sensor fabrication steps (with anodically grown WO_3 nanodot as sensing layer) is presented [22].

Based on the device structure, the response time and the recovery time are determined. In short, smaller transport path will result in faster response time and recovery time while increase in effective sensing surface (due to electrochemical means) will determine the sensitivity/response magnitude towards a particular target species. In this section, sensing performances of the anodically grown oxide layers, reported to sense different chemical vapors, will be discussed with emphasis on the surface morphology which are subdivided again in accordance with the sensor device structure employing that particular morphology.

The response magnitude is calculated as the relative change in corresponding electrical parameter (resistance, conductance, current or voltage) in target species with respect to the value of that parameter at inert ambient (nitrogen, argon or synthetic air). The response time/recovery time is calculated as the time taken to reach the 90 % of saturation response after target gas pulse is on/off, respectively.
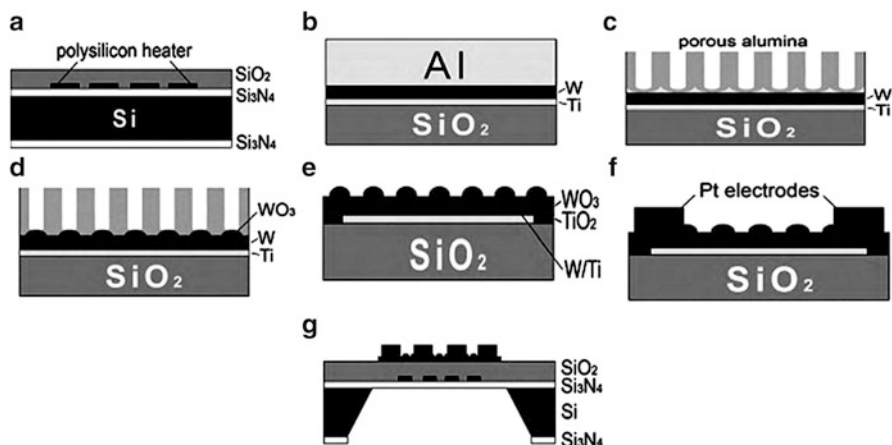
Fig. 14.7 Process steps for anodically grown tungsten oxide nanodot-based sensor device on micromachined platform: (**a**) Poly-Si heater and isolating layer (SiO_2) formation on processed Si wafer; (**b**) Sputtering of multilayer metal on this system; (**c–d**) Anodization process to form desired nanoforms; (**e**) removal of the top oxide layer to expose active sensing layer; (**f**) electrode formation for electrical characterization and (**g**) back etching of Si wafer (Reprinted with permission from [22])

3.1 Nanoporous

Different nanoporous semiconducting materials (TiO_2, ZnO, WO_3, SnO_2, Nb_2O_5), grown by electrochemical route, were reported for detection of a large variety of chemical species (H_2, NO_2, CO, CH_4, $HCHO$, CH_3COCH_3, NH_3 and H_2S) employing different device configurations.

3.1.1 Planar/Resistive

Using resistive device structure, a large number of publications were reported to sense hydrogen incorporating electrochemically grown nanoporous oxide as the sensing layer [24, 30, 33]. The sensing performance is highly dependable on the sensing material and its pore dimension and distribution. It is established that the response magnitude is inversely proportional to the pore diameters whereas the response time varies directly with the same, if all the other parameters (gas concentration, operating temperature, etc.) remain same [30]. To the contrary, smaller pore dimension (of TiO_2) results in lower but faster response towards higher concentration of H_2 as reported by Sadek et al. [33]. Such discrepancy was possibly due to the fact that the rutile TiO_2 is less sensitive than its anatase counterpart [30, 33]. Hence, for determining sensing performance, the crystalline nature is also equally important [30, 33]. It is relevant to mention that the UV-assisted etching of oxide results in better sensing ability of resistive device [24]. Among all the devices

containing electrochemical TiO_2, this device (which is based on anodically etched pre-existing oxide) showed the fastest response time (6 s) at 0.1 % hydrogen with the highest response magnitude (99 %). Apart from H_2 sensing, low concentration CO (30 ppm) sensing performance was also reported by the resistive device based on TiO_2 layer prepared by MAO technique [48]. This device showed a reasonable response time (125 s) and recovery time (292 s) towards 30 ppm CO at 350 °C.

Moreover, such nanoporous metal oxides are also capable of sensing the hazardous species, like methane [27], hydrogen sulfide [3] and ammonia [10]. Methane, the inflammable one, was reported to be sensed by electrochemically grown ZnO-based device in planar mode [27]. The sensor showed fast response time (∼9 s) with poor response magnitude (∼20 %) towards 0.1 % methane at 70 °C. Jeun and Hong [3] reported on CuO-loaded SnO_2 layer sensitive to very low concentration of hydrogen sulfide at 250 °C. For an exposure towards 5 ppm H_2S, the device detected the gas with ∼95 % response magnitude with a response time of ∼45 s and recovery time of ∼55 s. On the other hand, electrochemically derived porous alumina layer with better pore distribution improved response magnitude towards ammonia at room temperature (23 °C) [10]. High response magnitude was reported in case of porous tungsten oxide for sensing nitrogen dioxide at 150 °C [21, 34]. From these reports, it is confirmed that the sensing performance of porous oxide surface [21] is better than the cracked or the smooth ones [34]. It can be envisaged that, the smaller grain size [21] enhances adsorption rate of test vapor at even low concentration (1 ppm). It was also reported that, anodic WO_3 showed selectivity towards NO_2, at even lower concentrations, in comparison to100 ppm NH_3 and 500 ppm C_2H_5OH [21]. A representative transient response characteristic of resistive nanoporous anodic oxide layer is depicted in Fig. 14.8.

3.1.2 Schottky

The Schottky diode/junction is a special type of rectifying contact which is strongly depended on the type (work function) of the constituent metals and the semi-
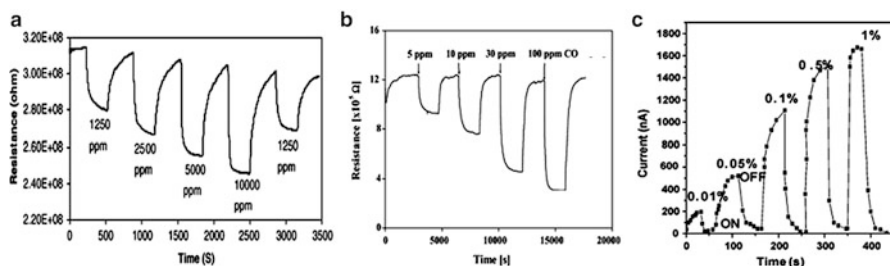
Fig. 14.8 Transient characteristics of resistive device towards different vapors, viz. (**a**) H_2, (**b**) CO, and (**c**) CH_4, incorporating nanoporous TiO_2 (**a**, **b**) and ZnO (**c**) layer at corresponding optimum temperature (reprinted with permission from [27, 33, 48])

conductors [4]. The device schematic is similar to that of the MIM configuration (Fig. 14.6b) where the bottom electrode serves as Ohmic contact and top electrode is responsible for schottky (rectifying) nature.

The gas sensors based on Schottky effect were reported in very few papers, where the sensor was made of electrochemically grown/etched oxide sensing layer [4]. The sensing performance of doped as well as undoped nanoporous layer were reported and compared for different electrodes (Pt, Pd–Ag, and Rh) [4]. However, Schottky junction gas sensors with electrochemically grown sensing layer is a relatively less investigated area.

3.1.3 MIM

In case of the MIM gas sensor device configuration, for the sensing purpose, the electrical measurements are usually carried out either in the constant dc voltage mode or in the constant dc current mode [24–29, 31, 32, 36–38]. These sensors have been developed to sense hydrogen [24–26, 28, 29, 31, 32] and methane [27, 36–38] using nanoporous metal oxide layer.

In case of hydrogen sensing, it can be envisaged that the porosity of the anodically grown film determines the sensing performance. Comparative study shows that the device, incorporating higher porosity titania film, offers higher response magnitude as well as faster response and recovery characteristics [25, 26, 31]. Iwanaga et al. [31] indicated that the application of reverse bias voltage for $Pd/TiO_2/Ti$ device results in higher sensitivity towards hydrogen. These effects can easily be understood from the data summarized in Table 14.2. Miyazaki et al. [32] reported on the effect of ambience on the devices based on oxide grown by using two different electrolytes (H_2SO_4 and HF). Metal oxide layer etched by UV-assisted anodization showed better sensitivity and faster response towards 500 and 1000 ppm H_2 than that of non-UV-assisted counterpart. Moreover, the latter one senses at higher temperature than that of the other [24]. It is reported that the effect of electrode dimension also influence the sensing properties in case of the MIM structure as well as in the resistive one (Fig. 14.9a, b).

Except ZnO and TiO_2, Nb_2O_5 was also reported to be sensitive towards hydrogen employing the MIM device configuration [28, 29]. From these studies, it can be pointed out that, for H_2 sensing, Pd electrode shows higher response magnitude than the devices with Au and Pt electrodes [29]. It was also confirmed that, the higher porosity sensing layer was responsible for faster response and recovery [28]. Sensing performance of MIM device towards hydrogen and methane is shown in Fig. 14.10.

Table 14.2 also reveals that nanoporous TiO_2, grown in H_2SO_4 medium, offered the highest sensitivity towards hydrogen [25]. Faster response was achieved by using Pd electrode [24] instead of Pt electrode [25]. Similarly, in case of methane sensing, low-temperature detection with appreciable response magnitude having fast response/recovery time was achieved by incorporating Pd-Ag electrode (on ZnO) and higher anodization voltage during preparation of sensing layer [27].

Table 14.2 Sensing performance of electrochemically derived oxide-based devices

Target species	Sensing material	Device structure	T (°C)	Sensing performance			C (ppm)	Reference
				RM	$\tau_1{}^a$	$\tau_2{}^a$		
H_2	Nanoporous TiO_2	Resistive	300	99 %	8.57 s	27.14 s	1000	[24]
			300	74 %	218 s	39 min	500	[30]
			225	19.35 %	23.81 s	158.71 s	10,000	[33]
		MIM	300	6	30.34 s	22.85 s	10,000	[25, 26]
			250	184	68.57 s	86 s	8000	[32]
			250	291.3	98 s	70.41 s	10,000	[31]
			300	98.79 %	6.25 s	177.14 s	1000	[24]
	Nanoporous Nb_2O_5	MIM	100	23.32	130 s	104 s	8000	[29]
			100	0.331	32.46 s	97.40 s	10,000	[28]
	TiO_2 nanotube	Resistive	25	99.97 %	14.28 s	82.85 s	500	[5]
			290	98.57 %	478 s	143 s	1000	[10]
		MIM	150	26.93	834 s	1544 s	1000	[11]
	WO_3 nanodots	MEMS	200	98.73 %	47.61 s	262 s	1000	[22]
NO_2	Nanoporous WO_3	Resistive	150	6.25	190 s	397 s	2	[34]
			150	41.2	101 s	93.17 s	1	[21]
	TiO_2 nanotube	Resistive	400	65 %	–	–	100	[7]
CO	Nanoporous TiO_2	Resistive	350	62.68 %	177 s	417 s	30	[48]
	Nanotube TiO_2	Resistive	400	23.92 %	388 s	500 s	75	[32]
	ZnO nanowires	Resistive	500	1.78	136 s	445 s	500	[20]
CH_4	Nanoporous ZnO	Resistive	70	39.3	13 s	28.57 s	1000	[27]
		Schottky	100	32.2	3.84 s	23 s	10,000	[4]
		MIM	220	3.85	5.1 s^b	16.1 s^b	10,000	[37]
			220	8.9	25 s	67.85 s	10,000	[36]
			240	22.6	12.53 s	36 s	10,000	[38]
			100	32.2	2.85 s	17 s	10,000	[27]

Gas	Material	Type	T	RM%	τ_1	τ_2	C	Ref
HCHO	TiO$_2$ nanotube	Resistive	19	97.62 %	44 s	571 s	50	[9]
	Nanoporous TiO$_2$		19	55.10 %	106 s	217.14 s	50	[8]
			22	36.36 %	–	–	50	[23]
CH$_3$COCH$_3$	ZnO nanowires	Resistive	500	146.06	343 s	300 s	100	[20]
	TiO$_2$ nanotube	MIM	200	~85 %	30 s	34 s	1000	[18]
	Pd-modified TiO$_2$ nanotube	Resistive	100	97 %	19 s	24 s	1000	[17]
NH$_3$	Nanoporous Al$_2$O$_3$	Resistive	23	47.58 %	214 s	357.14 s	100×10^4	[35]
	TiO$_2$ nanotube	Resistive	RT	4.3 %	105.67 s	48.97 s	150	[12]
			RT	34.55 %	308 s	26.78 s		
H$_2$S	Nanoporous SnO$_2$	Resistive	250	94.78 %	64.30 s	80.28 s	5	[3]
C$_2$H$_5$OH	TiO$_2$ nanotube	Resistive	RT	27.5 %	115.53 s	44.71 s	400	[12]
			200	35.54 %	21 min	11 min	5000	[13]
			RT	~55 %	~83 s	~330 s	700	[15]
C$_6$H$_6$		Resistive	150	48.72 %	56.13 s	152.96 s	400	[19]
C$_6$H$_5$CH$_3$		Resistive	150	41.62 %	84.93 s	128.54 s		
C$_6$H$_4$(CH$_3$)$_2$		Resistive	150	38.03 %	103.8 s	152.3 s		

RM response magnitude: ratio of change in resistance/conductance/impedance/current/voltage to the ambient value of corresponding parameters (optionally expressed in percentage), *T* operating temperature for sensing, *C* concentration of vapor, *RT* room temperature

[a] Response time (τ_1) and recovery time (τ_2) is calculated based on the time taken by sensor to reach its 90 % of saturation and baseline values respectively

[b] Response time and recovery time was calculated based on the time taken by sensor to reach its 67 % of saturation and baseline values respectively
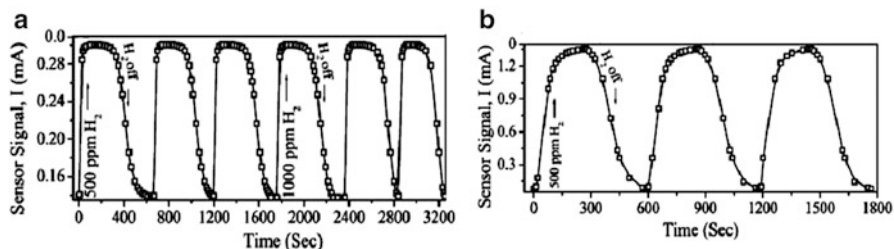
Fig. 14.9 Effect of electrode (Pd) dimension: (**a**) 3 mm and (**b**) 2 mm in radius, towards H_2 with nanoporous TiO_2-based MIM device at 250 °C (reprinted with permission from [24])
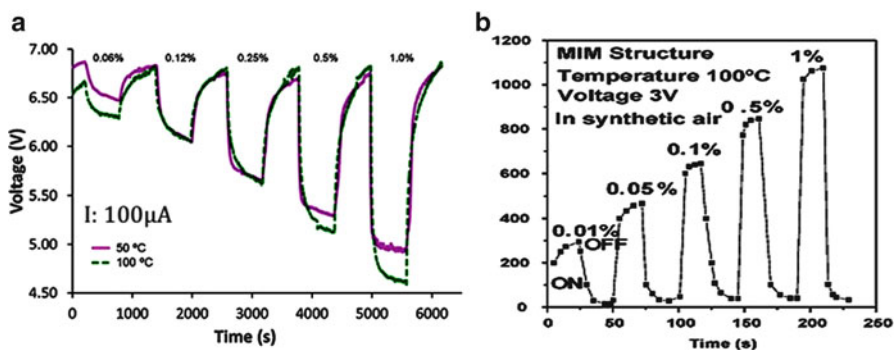
Fig. 14.10 Transient response of MIM-based device towards (**a**) 0.06–1 % H_2 by nanoporous Nb_2O_5 layer and (**b**) 0.01–1 % CH_4 by nanoporous ZnO layer (Reprinted with permission from [27, 28])

3.2 Nanotubes

Due to lengthwise distribution of active adsorption sites on both outer and inner walls, nanotube possesses higher surface area than that of ordinary nanoporous structure. Electrochemical anodization is perhaps the most controllable route to grow tubes having length and diameter in the nanometer range. Different parameters of anodization have immense effect on tube length, diameter and wall thickness. It is worth mentioning that, no other oxides nanotubes (NT) other than TiO_2, derived through electrochemical route have been reported so far for gas sensing [5–19]. TiO_2 NT-based sensors are sensitive to a large variety of target species including H_2 [5, 10, 11], NO_2 [6, 7], CO [7], HCHO [8, 9], NH_3 [12], C_2H_5OH [12–16], CCl_4 [12], and acetone [17, 18]. Additionally, detection of some aromatic hydrocarbons (e.g., benzene, toluene and xylene or BTX) was also reported incorporating such nanotubular configuration [19]. The sensing ability of nanotube-based sensors is discussed for different device configuration in next subsections.

3.2.1 Resistive

It has been experimentally authenticated that, the longer nanotubes with shorter diameters result in higher response magnitude and faster response/recovery characteristics, for H_2 and NO_2 detection [5–7, 10, 11] as reflected from the data summarized in Table 14.2. Aiming towards larger surface to volume ratio, the nanotubular film was pulverized, by ball-milling method, which facilitated better sensing performance towards lower concentration of species (formaldehyde) as the gas interaction sites became larger [8, 9]. Humid ambient facilitated the sensing performance but deteriorated the recovery process for formaldehyde sensing purpose. Wu et al. [8] also reported that the sensor showed greater sensitivity and stable baseline resistance if UV ray was irradiated at the time of sensing (HCHO).

The effect of relative humidity on ammonia and ethanol sensing by TiO_2 nanotubular layer was also reported by Perillo and Rodríguez [12]. Here, nanotubular structure showed faster response as well as recovery with higher response magnitude towards 400 ppm ethanol in case of low relative humidity; but the lower humidity showed slower response for ammonia (150 ppm) sensing, keeping other parameter same. Kılınç et al. [13] reported that the sensitivity towards different VOCs (5000 ppm of ethanol, dichloromethane, chloroform, carbon tetrachloride, methanol and isopropyl alcohol) at 200 °C was better for the device with anodic TiO_2 prepared in aqueous HF electrolyte than the TiO_2 layer prepared in EG-based NH_4F electrolyte, due to their respective morphological aspects. A noticeable fact is that, the device with anodic TiO_2 prepared in aqueous HF electrolyte showed faster recovery than response in case of C_2H_5OH sensing and in contrast, the same device showed more recovery time than response time in case of CCl_4 sensing.

On the other hand, Hazra et al. [15] and Dutta et al. [19] described that the amount of water within viscous electrolyte, during electrochemical synthesis, determined the sensing performance of the titania nanotubular film towards alcohols and BTX by optimizing surface to volume ratio and intrinsic oxygen vacancies. The transient characteristic of the optimized TiO_2 nanotube array, in terms of oxygen vacancy and surface to volume ratio, is shown in Fig. 14.11. Based on such optimized condition of electrolyte, effect of variation of length and wall thickness of nanotube on alcohol sensing was investigated in details by Hazra et al. [16]. It is observed that, variation of the tube length was not so effective whereas, a reduction in wall thickness resulted in higher sensitivity. Bhattacharyya et al. [17] reported on the effect of catalytic metal (Ni and Pd) modification on acetone sensing performance of TiO_2 nanotubes, where it was concluded that modification is a route for betterment of response magnitude and modification with Pd nanoparticle is better than that of Ni counterpart.

3.2.2 MIM

To achieve faster response (than that of resistive one), the MIM structure was introduced for metal oxide-based chemical sensor. Hydrogen sensing performance
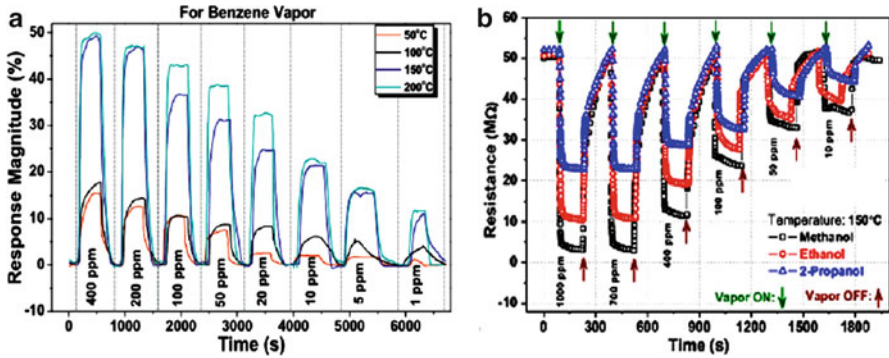
Fig. 14.11 Sensing performance of TiO$_2$ nanotube array-based resistive device towards (**a**) 1–400 ppm benzene and (**b**) 10–1000 ppm of different alcohols (reprinted with permission from [16, 19])
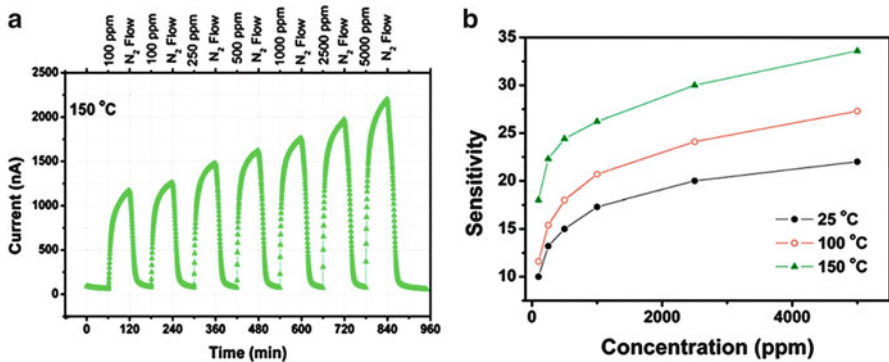
Fig. 14.12 Sensing performance of nanotubular TiO$_2$-based MIM device towards hydrogen: (**a**) Transient response at 150 °C and (**b**) variation of sensitivity with change in gas concentration at different temperatures (reprinted with permission from [11])

of the device (Fig. 14.12), based on nanotubular TiO$_2$ layer, operating in MIM configuration was reported by Şennik et al. [11]. At 150 °C, the device sensed 1000 ppm H$_2$ with ∼27 % response magnitude and 600 s response time. It was pointed out that, at lower temperature (25 °C) the base line drift was observed for transient characteristics (H$_2$ sensing) whereas this drift was absent for higher temperature (150 °C). A detailed study on effect of length variation on acetone sensing was investigated by Bhowmik et al. [18]. It was revealed that the response magnitude increased with increase in tube length. On the other hand, increase in tube length resulted in elongated electron transport path for which response/recovery characteristics got slower in such vertical device configuration [18].

Comparison between titania nanotube-based planar and vertical device reveals that the latter one is low-power consuming device configuration as optimum sensing

temperature and actuating voltage is also lower towards ethanol sensing [15]. Response/recovery characteristic is faster for MIM configuration but at the cost of lower response magnitude.

3.3   Nanowires

Electrochemical anodization sometimes results in wire like structure of metal oxides with appreciably high surface to volume ratio. This structure is well separated and loosely connected to bulk.

The devices with ZnO nanowires [20] and WO_3 nanowires [55] were investigated in resistive configuration to sense CO, CH_3COCH_3, H_2, O_2, and NO. It was reported that ZnO nanowire-based device is more sensitive to 100 ppm acetone than 500 ppm CO at 500 °C [20]. Kukkola et al. [55] presented a review on anodic tungsten oxide with same surface morphologies targeting different species. It was revealed that 10 ppm of CO, H_2, and O_2 and 2 ppm of NO sensing was possible with acceptable response magnitude. CO was sensed with synthetic air and argon (separately) as the carrier at 200 °C. The response magnitude was more in case of argon than that in the air. The highest response towards 40 ppm CO in air ambient was achieved at 120 °C with unacceptably long recovery of almost 10 min. It was further revealed that the argon ambient helped the device to be more sensitive to a particular H_2 concentration than for the air, but in expense of slower recovery. However, the device offered different sensing characteristics towards NO when the test gas was tested with argon and air background, separately. For argon the sensor conductance decreased while the same increased for air. Oxidizing nature of air is probably responsible for such behavior. 170 °C was reported to be the optimum temperature for the highest response for sensing 40 ppm of O_2. In Fig. 14.13, the gas sensing properties of the device is presented for a better understanding of this discussion.

3.4   Nanodots

Apart from pores, tubes and wires, discrete nanodots of metal oxide were also investigated for potential chemical sensor applications [22]. The nanodots were developed electrochemically through the pores of anodic alumina patterning/template.

The sensing performance of the device, based on WO_3 nanodots, towards hydrogen was reported by Calavia et al. [22]. Incorporation of the microheater on the same platform facilitated the tuning of temperature-dependent sensitivity and selectivity. The response magnitude was reported to be 98.73 % towards 1000 ppm H_2 at 200 °C. Temperature dependence of response magnitude and the microscopic view of the device are shown in Fig. 14.14a, b.
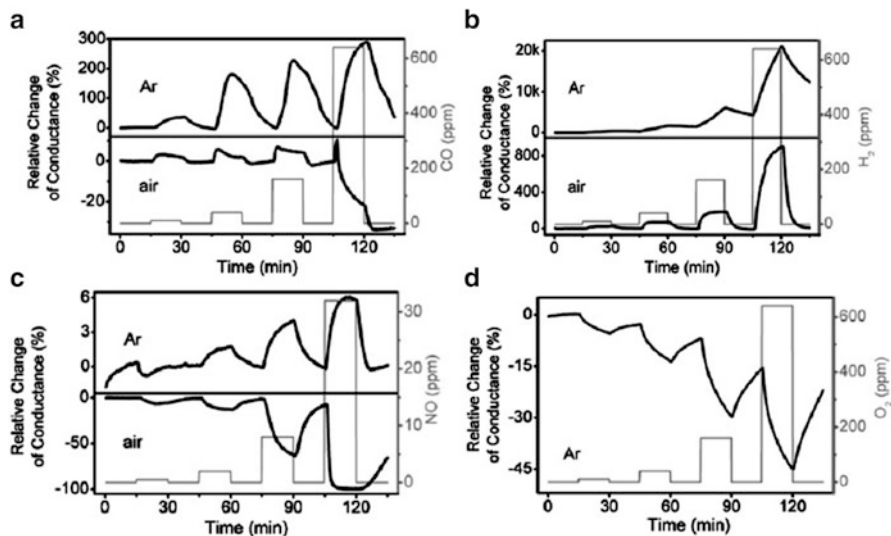
Fig. 14.13 Sensing performance of WO₃ nanowire-based sensor towards (**a**) CO, (**b**) H₂, (**c**) NO, and (**d**) O₂ in air and argon as carrier gas (reprinted with permission from [55])
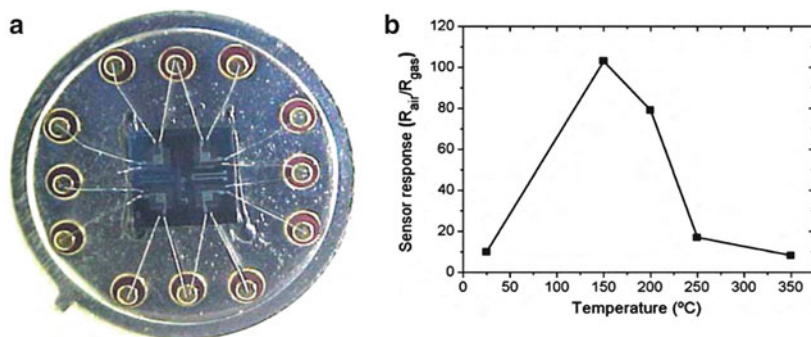
Fig. 14.14 WO₃ nanodot-based device (**a**) Optical microscope image of the sensor microchip and (**b**) its sensing performance towards 1000 ppm H₂ at various operating temperatures (reprinted with permission from [22])

A literature review on the electrochemically derived metal oxide-based chemical sensor has been presented in this section with some illustrative sensing characteristics and corresponding surface morphology.

From Table 14.2, it can be noticed that devices with WO₃ layer sensed low concentration of NO₂ at relatively lower operating temperature whereas devices with TiO₂ sensing layer offered higher operating temperature for sensing higher concentration of the same species [7, 21, 34]. Moreover, grain size plays a significant role as lower energy (i.e., lower thermal energy) is required for sensing same target molecules for the sensing layer having lower dimension of grains [58]. Lower

grain size of WO_3, than that of TiO_2, in the above case authenticated the above statement [7, 21, 34]. Again, comparing the sensing performance of TiO_2 and WO_3 towards hydrogen, latter offered lower response magnitude and slower response [8, 22]. Probably, discontinuous layer of WO_3 (in form of discrete nanodots) caused longer electron transport path (effective) which lead to sluggish response/recovery characteristics. For CO sensing performance, it was reported that ZnO responded more rapidly than TiO_2 towards the target gas, but at the cost of higher lowest detection level [20, 32, 48]. Use of catalytic metals as electrode improves the response time for ZnO-based devices. Lower concentration of ammonia was sensed by TiO_2 nanotubular layer with faster recovery, in comparison to Al_2O_3 [12, 35]. Possibly, higher surface area of TiO_2 nanotube provided higher amount of interaction sites than that of nanoporous alumina. Thus it can be observed that, salient features of nanostructuring (tuned by the parameters of electrochemical anodization) eventually determine the resultant sensing performance.

4 Integration with MEMS: Challenges and Prospects

4.1 Purpose, Schemes, and Scientific Issues Concerning Integration

For reliable sensor device performance, it is essential that, the optimum operating temperature for sensing should be maintained throughout the testing in presence of a particular species. To achieve this, microheater embedded gas sensor system is an effective means where apart from micro heater, the signal conditioning circuitry may also be integrated on the same platform [59]. Micromachining technique offers such type of heater fabrication on a thin membrane and additionally provides common platform for signal processing unit (such as dc voltage/current source, conditional amplifier, LED, alarm etc.).

Incorporation of metal oxide-based sensor devices on micromachined membrane have been extensively reviewed with detailed process steps of different approaches, except electrochemical anodization by Simon et al. [60] and Bhattacharyya [57]. It has been demonstrated that, the micromachining of Si substrate is prior step for deposition of sensing (metal oxide) layer by either liquid deposition technique or chemical/physical vapor deposition technique. Depending upon the micromachining technique, desired membrane structure is achieved either by front side etching (suspended membrane) or by backside etching (closed membrane). The difference of these two approaches has been elucidated in aforesaid reviews [57, 60].

Bhattacharyya [57] further described the present technological scenario of microheater fabrication where it was pointed out that, the wet etching (suitable for bulk micromachining) is usually avoided in sophisticated minimized devices as the dry etching offers higher resolution than that of wet one. In this regard, sacrificial layer of SiO_2 or phosposilicate glass is used for selective etching purpose

(surface micromachining technique) [57]. Besides, a detailed understanding of heat transfer from the microheater to the sensing layer and to the ambience, is also a pivotal factor for successful device performance [60]. However, in case of electrochemically grown metal oxides (as sensing layer), some deviations are encountered while incorporating this with micromachining (MEMS) technology. In most of the cases, for synthesizing oxide nanoforms through anodization, different metals are used as the substrate [5–21, 23, 28]. However, standard micromachining technique, employed for MEMS fabrication, deals with silicon substrate and as a result of such substrate a metal layer needs to be pre-deposited on Si/SiO$_2$ substrate for subsequent anodization. Usually, thermal evaporation, e-beam evaporation sputtering etc. are employed to deposit such metals [3, 20, 22]. The challenges and prospects of integrating such electrochemically grown oxide nanoforms with standard micromachined substrate have been discussed in the subsequent sections.

4.2 Challenges

4.2.1 Selective Window for Anodization

As already discussed in the earlier sections, the anodization procedure must be carried out after deposition of the metal layer on the substrate/membrane and electrochemical process may jeopardize the substrate/membrane structure. On the other hand, sputtering on membrane (achieved after micromachining) is disadvantageous as particles with high kinetic energy at higher temperature may cause mechanical fracture in the membrane [61]. Therefore, to prevent any mechanical stress, during metal deposition, micromachining is carried out at the last stage of the fabrication. Moreover, during electrochemical anodization the electrolyte should be chosen carefully such that, it should not chemically react with the other parts of the device. To achieve this, usually, a protective layer (usually PTFE [22]) is used to ensure a selective window where the metal needs to be oxidized. Metal masking is not possible in such cases. But, it should also be ensured that, removal of the protecting barrier should be easy and that must not indulge any physical or chemical change of the sensing layer. In this regard, Calavia et al. [22] demonstrated the usefulness of polytetrafluoroethylene (PTFE) as the masking material, which provides isolation of Si wafer (and related layers like SiO$_2$, Si$_3$N$_4$, etc.) from the anodizing electrolyte and oxidizes the underlying metal through a circular window on metal layer.

4.2.2 Metal Thickness and Adhesion

It is already pointed out that the anodizing metal should be deposited on Si/SiO$_2$ wafer prior to anodization. In such cases, metal thickness should be critically considered while optimizing the anodization parameters. As per earlier discussions, the thickness of the nanotubes/nanorods depends on the anodization time, type and

constituents of electrolyte etc. Therefore, selection of these parameters as well as final tube/rod lengths depends on the metal layer thickness. In the worst case, the electrolyte may etch/anodize the whole metal layer thickness and the sensing layer (metal oxide) may no longer adhere to the substrate. In turn, the use of selective anodization window fails to serve its purpose. Moreover, even if sufficiently thick metal layer is deposited on the substrate with poor adhesion (to the substrate), there is also a chance that the sensing layer (after anodization) will not adhere to MEMS platform. Hence, metal adhesion on substrate (Si or SiO$_2$ or other insulating layer) is of prime concern, beside the thickness. To overcome this, controlled substrate heating, during metal deposition, is one of the way which ensures better adhesion of metal layer on the substrate [62]. In another approach, adhesive metal (Ti, Cr, etc.) layer(s) (comparatively thinner) is pre-deposited on the substrate and subsequently metal layer (to be anodized) is deposited with relatively better adhesion [22].

4.2.3   Selective Window for Metallization

For MEMS-based gas sensor development, with electrochemically anodized metal oxides as the sensing layer, metallizations at three different stages are required, viz. (i) for microheater fabrication, (ii) contact electrode fabrication and (iii) as a base material of anodized oxide-sensing layer synthesis. Depending upon the device geometry, the microheater and electrodes can either be situated on the same layer (horizontal approach) or on the different layers (vertical approach) [57, 60]. However, it is to be ensured that there is no unwanted short circuit of these metal lines deposited for the above three purposes. To ensure this, usually, these metals are also deposited in a selective manner using shadow metal mask during metal deposition. Further, the process becomes relatively simplified (from metallization perspective) if the polysilicon heater is employed in place of the metal heater.

4.2.4   Selective Window for Si Etching

The membrane formation is accomplished after the anodization process and here also, a protective layer should be provided relinquishing the probability of unwanted etching of the undesired part. If the etching is carried out from the backside of the wafer, the front side nanostructured sensing layer has to be protected from the Si etchants. During the protective masking, retaining the nanoform (structural and electronic properties) under the coverage (and subsequent removal) of protective layer is a crucial challenge. Clearly, for surface micromachining or front side etching, such protective layer employment is even more challenging.

4.2.5 Dry Etching Preference

Etching of the Si substrate to produce the membrane is one of the most tricky fabrication steps in micromachining system where the membrane allows easy and low-loss transportation of heat to the active (sensing) area. Etching of silicon substrate is categorized as isotropic and anisotropic one, depending upon the directionality of etching [59]. In short, the isotropic etching is responsible for semicircular etching whereas anisotropic etching is the basis of vertical etching (high-aspect ratio). On the other hand, the etching technique can further be categorized into dry etching and wet etching, depending on the type of etchant [59]. In general, wet etching is usually an isotropic one and the etch rate depends on the materials to be etched and physical parameters like temperature. Potassium hydroxide (KOH) ethylene diaminepyrochatechol (EDP) and tetramethyl ammonium hydroxide (TMAH) is used as wet etchant which etches according to crystallographic direction [59, 60] and popular etch stops are SiO_2 or Si_3N_4 for this purpose [60]. But, in case with anodized oxide sensing layer, these etchants may be chemically corrosive to the oxide layer (sensing) which may jeopardize the sensing properties significantly.

To the contrary, dry etching is featured with smaller undercut and higher anisotropy but compromises with the material selectivity, in comparison to its wet counterpart [59]. Depending on etching mechanism, physical dry etching (ion milling) etches with low etch rate with chemically inert gas (argon) whereas, chemical dry etching (high pressure plasma) is used with highly reactive species (XeF_2) improving the etch rate substantially [59]. Combination of the physical and chemical method called reactive ion etching (RIE) is very popular and well-equipped means in modern MEMS technology [57]. Anisotropy may be enhanced through deep reactive ion etching (DRIE) [57, 59]. Protection of sensing layer is necessary as dry etching is generally performed from the front side. RIE and DRIE automatically allow such masking as the ion activated area needs to be predefined prior to etching [59]. Hence, it will be suitable for anodized oxide-based sensor device with micromachined Si substrate. Calavia et al. [22] successfully demonstrated micromachined platform having anodized WO_3 nanodots on the top of the membrane employing wet etching technique. Most possibly, selective etching property of wet etchant (KOH for the case), in their case, does not affect the oxide layer chemically or physically as the process ensures lower temperature (<60 °C). Thus, if such etchants can be identified for individual of anodized oxide sensing layer and bulk micromachining is not an issue of importance; then wet etching can also be employed for anodized film-based gas sensor development with MEMS platform.

4.3 Prospects

It can be envisaged that the challenges associated in integrating the electrochemically anodized nanostructured sensing layer with the micromachined silicon substrate calls for multidimensional optimization of process steps, process

parameters and process reagents. For selective processing at different levels, mask design becomes complicated and total number of process steps involved also increases significantly. In spite of that, due to unprecedented advantages of the electrochemically grown nanostructured sensing layer, modern chemical sensor researchers attempted to integrate the same with MEMS platform [22]. Complexity regarding them at different stages of metallization and subsequent etching can be partially circumvented by employing polysilicon heater in place of metal ones [57]. On the other hand, for better adhesion of the metal layers with desired thickness (for electrode and base metal for anodization) a few nanometer thick Ti/Cr adhesive layer is usually pre-deposited [22, 57]. Selective anodization is carried out such that the sensing layer is situated at the center of the active area/membrane. For such anodization, often aluminum template is used. PTFE or similar materials are usually employed to protect the other part of device during anodization [22]. Micromachining is carried out at the last step of process flow and temperature of etchant is precisely maintained so that it does not react with anodized sensing layer. Moreover, judicious selection of the dry/wet etchants is also necessary.

5 Concluding Remarks

A critical and thorough review on vapor/gas sensing properties of a wide range of electrochemically derived metal oxide nanoforms as the sensing layer employing different device configuration has been conducted in this chapter. The challenges, prospects, and possible measures to mitigate the limiting issues (for integrate electrochemically derived sensing layer with MEMS platform) have been elucidated. An approach to understand the mechanism behind the oxide growth has been represented followed by the effect of anodization parameters on oxide properties. Variations of electrochemical parameters (like anodization time, temperature, applied electric field, constituents and concentrations of electrolyte etc.) eventually control the morphology as well as the intrinsic properties of the grown oxide(s). It is also observed that, depending on the native metal, the porosity and the dimension of the nanoforms varies even if other anodization parameters remain unchanged. Besides, crystalline phase, grain size, and stoichiometry of metal oxides directly influence the sensing performance of the device. Considering the sensing performance of all the nanoforms, it is observed that, nanotube array, due to their highest specific surface area is capable of sensing the target species with better efficiency than that of the other nanoforms. Literature review suggests that, different device configuration has been incorporated among which MIM device structure is a promising one as it offers faster response/recovery compared to its conventional resistive counterpart. On the other hand, micromachined platform is suitable for in-built low-power heater assembly which maintains the optimum operating temperature at the test field, along with the chance to integrate the supporting electronic circuitry on the same platform. But, such integration of anodized sensing layer with the micromachined platform involves a large number of process steps which must be

dealt with proper expertise. These additional steps increase the cost of final product as well as complexity, but with better device performance (e.g., better sensitivity, speed, and power). Therefore, successful attempts are made recently to integrate electrochemically grown efficient sensing layer with the efficient platform fabricated through micromachining.

References

1. Fine GF, Cavanagh LM, Afonja A, Binions R (2010) Metal oxide semi-conductor gas sensors in environmental monitoring. Sensors 10:5469–5502
2. Wang C, Yin L, Zhang L, Xiang D, Gao R (2010) Metal oxide gas sensors: sensitivity and influencing factors. Sensors 10:2088–2106
3. Jeun J, Hong S (2010) CuO-loaded nano-porous SnO_2 films fabricated by anodic oxidation and RIE process and their gas sensing properties. Sens Actuators B Chem 151:1–7
4. Basu PK, Saha N, Jana SK, Saha H, Spetz AL, Basu S (2009) Schottky junction methane sensors using electrochemically grown nanocrystalline-nanoporous ZnO thin films. J. Sens 790(476). doi: 10.1155/2009/790476
5. Chen K, Xie K, Feng X, Wang S, Hu R, Gu H, Li Y (2012) An excellent room-temperature hydrogen sensor based on titania nanotube-arrays. Int J Hydrogen Energy 37:13602–13609
6. Gönüllü Y, Rodríguez CGM, Saruhan B, Bayata F, Ürgen M (2011) Effect of doping at TiO_2-nanotubular gas sensors. Sens Proc D2(4):551–553
7. Gönüllü Y, César G, Rodríguez M, Saruhan B, Ürgen M (2012) Improvement of gas sensing performance of TiO_2 towards NO_2 by nano-tubular structuring. Sens Actuators B Chem 169:151–160
8. Wu G, Zhang J, Wang X, Liao J, Xia H, Akbar SA, Li J, Lin S, Li X, Wang J (2012) Hierarchical structured TiO_2 nano-tubes for formaldehyde sensing. Ceram Int 38:6341–6347
9. Lin S, Li D, Wu J, Li X, Akbar SA (2011) A selective room temperature formaldehyde gas sensor using TiO_2 nanotube arrays. Sens Actuators B Chem 156:505–509
10. Varghese OK, Gong D, Paulose M, Ong KG, Grimes CA (2003) Hydrogen sensing using titania nanotubes. Sens Actuators B Chem 93:338–344
11. Şennik E, Çolak Z, Kılınç N, Öztürk ZZ (2010) Synthesis of highly-ordered TiO_2 nanotubes for a hydrogen sensor. Int J Hydrogen Energy 35:4420–4427
12. Perillo PM, Rodríguez DF (2012) The gas sensing properties at room temperature of TiO_2 nanotubes by anodization. Sens Actuators B Chem 171–172:639–643
13. Kılınç N, Şennik E, Öztürk ZZ (2011) Fabrication of TiO_2 nanotubes by anodization of Ti thin films for VOC sensing. Thin Solid Films 520:953–958
14. Hazra A, Bhattacharyya P (2014) Tailoring of the gas sensing performance of TiO_2 nanotubes by 1-D vertical electron. IEEE Trans Electron Devices 61:3483–3489
15. Hazra A, Dutta K, Bhowmik B, Chattopadhyay PP, Bhattacharyya P (2014) Room temperature alcohol sensing by oxygen vacancy controlled TiO_2 nanotube array. Appl Phys Lett 105:081604
16. Hazra A, Bhowmik B, Dutta K, Chattopadhyay PP, Bhattacharyya P (2015) Stoichiometry, length, and wall thickness optimization of TiO_2 nanotube array for efficient alcohol sensing. Appl Mater Interfaces 7:9336–9348
17. Bhattacharyya P, Bhowmik B, Fecht H (2015) Operating temperature, repeatability, and selectivity of TiO_2 nanotube-based acetone sensor: influence of Pd and Ni nanoparticle modifications. IEEE Trans Device Mater Reliab 15:376–383
18. Bhowmik B, Fecht H, Bhattacharyya P (2015) Vertical mode gas sensing performance of TiO_2 nanotube array by tuning of surface area and carrier transport length. IEEE Sens 15:5919–5926

19. Dutta K, Chattopadhyay PP, Lu CW, Ho MS, Bhattacharyya P (2015) A highly sensitive BTX sensor based on electrochemically derived wall connected TiO_2 nanotubes. Appl Surf Sci 354:353–361
20. Galstyan V, Comini E, Baratto C, Fagila G, Brisotto M, Bontempi E, Sberveglien G (2012) Growth and gas sensing properties of rough ZnO nanowires. IMCS, pp 1623–1625
21. Zeng J, Hu M, Wang W, Chen H, Qin Y (2012) NO_2-sensing properties of porous WO_3 gas sensor based on anodized sputtered tungsten thin film. Sens Actuators B Chem 161:447–452
22. Calavia R, Mozalev A, Vazquez R, Gracia I, Cané C, Ionescu R, Llobet E (2010) Fabrication of WO_3 nanodot-based microsensors highly sensitive to hydrogen. Sens Actuators B Chem 149:352–361
23. Liu L, Li X, Wang X, Li S, Liao J, Tang Z, Wang J (2012) UV assisted chemical gas sensing of nanoporous TiO_2 at low temperature. IMCS, pp 698–701
24. Hazra SK, Basu S (2006) High sensitivity and fast response hydrogen sensors based on electrochemically etched porous titania thin films. Sens Actuators B Chem 115:403–411
25. Shimizu Y, Kuwano N, Hyodo T, Egashira M (2002) High H_2 sensing performance of anodically oxidized TiO_2 film contacted with Pd. Sens Actuators B Chem 83:195–201
26. Shimizu Y, Hyodo T, Egashira M (2007) H_2 sensing performance of anodically oxidized TiO_2 thin films equipped with Pd electrode. Sens Actuators B Chem 121:219–230
27. Basu PK, Jana SK, Saha H, Basu S (2008) Low temperature methane sensing by electrochemically grown and surface modified ZnO thin films. Sens Actuators B Chem 135:81–88
28. Rani RA, Zoolfakar AS, Ou JZ, Field MR, Austin M, Kalantar-zadeh K (2013) Nanoporous Nb_2O_5 hydrogen gas sensor. Sens Actuators B Chem 176:149–156
29. Hyodo T, Ohoka J, Shimizu Y, Egashira M (2006) Design of anodically oxidized Nb_2O_5 films as a diode-type H_2 sensing material. Sens Actuators B Chem 117:359–366
30. Mukherjee T, Hazra SK, Basu S (2007) Porous titania thin films grown by anodic oxidation for hydrogen sensors porous titania thin films grown by anodic oxidation. Mater Manuf Processes 21:247–251
31. Iwanaga T, Hyodo T, Shimizu Y, Egashira M (2003) H_2 sensing properties and mechanism of anodically oxidized TiO_2 film contacted with Pd electrode. Sens Actuators B Chem 93:519–525
32. Miyazaki H, Hyodo T, Shimizu Y, Egashira M (2005) Hydrogen-sensing properties of anodically oxidized TiO_2 film sensors: effects of preparation and pretreatment conditions. Sens Actuators B Chem 108:467–472
33. Sadek AZ, Partridge JG, McCulloch DG, Li YX, Yu XF, Wlodarski W, Kalantar-zadeh K (2009) Nanoporous TiO_2 thin film based conductometric H_2 sensor. Thin Solid Films 518:1294–1298
34. Hu M, Zeng J, Wang W, Chen H, Qin Y (2011) Porous WO_3 from anodized sputtered tungsten thin films for NO_2 detection. Appl Surf Sci 258:1062–1068
35. Varghese OK, Gong D, Dreschel WR, Ong KG, Grimes CA (2003) Ammonia detection using nanoporous alumina resistive and surface acoustic wave sensors. Sens Actuators B Chem 94:27–35
36. Basu PK, Bhattacharyya P, Saha N, Saha H, Basu S (2008) The superior performance of the electrochemically grown ZnO thin films as methane sensor. Sens Actuators B Chem 133:357–363
37. Bhattacharyya P, Basu PK, Basu S (2011) Methane detection by nano ZnO based MIM sensor devices. Sens Transducers 10:121–130
38. Basu PK, Bhattacharyya P, Saha N, Saha H, Basu S (2008) Methane sensing properties of platinum catalysed nano porous zinc oxide thin films derived by electrochemical anodization. Sens Lett 6:1–7
39. Macak JM, Schmuki P (2006) Anodic growth of self-organized anodic TiO_2 nanotubes in viscous electrolytes. Electrochim Acta 52:1258–1264
40. Hazra A, Bhowmik B, Dutta K, Manjuladevi V, Gupta RK, Chattopadhyay PP, Bhattacharyya P (2014) Formation mechanism of anodically grown free-standing TiO_2 nanotube array under the influence of mixed electrolytes. Sci Adv Mater 6:714–719

41. Yang H, Pan C (2010) Diameter-controlled growth of TiO$_2$ nanotube arrays by anodization and its photoelectric property. J Alloys Compd 492:2009–2011

42. Macak JM, Tsuchiya H, Ghicov A, Yasuda K, Hahn R, Bauer S, Schmuki P (2007) TiO$_2$ nanotubes: self-organized electrochemical formation, properties and applications. Curr Opin Solid State Mater Sci 11:3–18

43. Liu G, Wang K, Hoivik N, Jakobsen H (2012) Solar energy materials & solar cells progress on free-standing and flow-through TiO$_2$ nanotube membranes. Sol Energy Mater Sol Cells 98:24–38

44. Choi J, Wehrspohn RB, Lee J, Gösele U (2004) Anodization of nanoimprinted titanium: a comparison with formation of porous alumina. Electrochim Acta 49:2645–2652

45. Jackson MJ (2007) Surface engineered surgical tools and medical devices surface engineered surgical tools and medical devices. Springer, Berlin. e-ISBN: 978-0-387-27028-9

46. Anitha VC, Menon D, Nair SV, Prasanth R (2010) Electrochemical tuning of titania nanotube morphology in inhibitor electrolytes. Electrochim Acta 55:3703–3713

47. Sugiura T, Yoshida T, Minoura H (1998) Designing a TiO$_2$ nano-honeycomb structure using photoelectrochemical etching. Electrochem Solid State Lett 1:175–177

48. Jun YK, Kim HS, Lee JH, Hong SH (2006) CO sensing performance in micro-arc oxidized TiO$_2$ films for air quality control. Sens Actuators B Chem 120:69–73

49. Tsuchiya H, Schmuki P (2005) Self-organized high aspect ratio porous hafnium oxide prepared by electrochemical anodization. Electrochem Commun 7:49–52

50. Qiu X, Howe JY, Meyer HM, Tuncer E, Paranthaman MP (2011) Thermal stability of HfO$_2$ nanotube arrays. Appl Surf Sci 257:4075–4081

51. Fang D, Yu JG, Luo ZP, Liu S, Huang K, Xu W (2012) Fabrication parameter-dependent morphologies of self-organized ZrO$_2$ nanotubes during anodization. J Solid State Electrochem 16:1219–1228

52. Hazra A, Dutta K, Bhowmik B, Manjuladevi V, Gupta RK, Chattopadhyay PP, Bhattacharyya P (2014) Structural and optical characterizations of electrochemically grown connected and free-standing TiO$_2$ nanotube array. J Electron Mater 43:3229–3235

53. Chen X, Schriver M, Suen T, Mao SS (2007) Fabrication of 10 nm diameter TiO$_2$ nanotube arrays by titanium anodization. Thin Solid Films 515:8511–8514

54. Allam NK, Shankar K, Grimes CA (2008) Photoelectrochemical and water photoelectrolysis properties of ordered TiO$_2$ nanotubes fabricated by Ti anodization in fluoride-free HCl electrolytes. J Mater Chem 18:2341–2348

55. Kukkola J, Mäklin J, Halonen N, Kyllönen T, Tóth G, Szabó M, Shchukarev A, Mikkola JP, Jantunen H, Kordás K (2011) Gas sensors based on anodic tungsten oxide. Sens Actuators B Chem 153:293–300

56. Sastry NV, Patel MC (2003) Densities, excess molar volumes, viscosities, speeds of sound, excess isentropic compressibilities, and relative permittivities for alkyl (methyl, ethyl, butyl, and isoamyl) acetates + glycols at different temperatures. J Chem Eng Data 48:1019–1027

57. Bhattacharyya P (2014) Technological journey towards reliable microheater development for MEMS gas sensors: a review. IEEE Trans Device Mater Reliab 14:589–599

58. Gurlo A, Ivanovskaya M, Bârsan N, Schweizer-Berberich M, Weimar U, Göpel W, Diéguez A (1997) Grain size control in nanocrystalline In$_2$O$_3$ semiconductor gas sensors. Sens Actuators B Chem 44:327–333

59. Ziaie B, Baldi A, Lei M, Gu Y, Siegel RA (2004) Hard and soft micromachining for BioMEMS: review of techniques and examples of applications in microfluidics and drug delivery. Adv Drug Deliv Rev 56:145–172

60. Simon I, Bârsan N, Bauer M, Weimar U (2001) Micromachined metal oxide gas sensors: opportunities to improve sensor performance. Sens Actuators B Chem 73:1–26

61. Sigmund P (1969) Theory of sputtering. Phys Rev 184:383–416

62. Zhang G, Zhao J, Green MA (1998) Effect of substrate heating on the adhesion and humidity resistance of evaporated MgF$_2$/ZnS antireflection coatings and on the performance of high-efficiency silicon solar cells. Sol Energy Mater Sol Cells 51:393–400

Chapter 15
Minimally Invasive Medical Devices and Healthcare Devices Using Microfabrication Technology

Y. Haga, T. Matsunaga, T. Kobayashi, and N. Tsuruoka

Abstract Medical tools for use in the human body, such as catheters and endoscopic tools, usually need to be thin and small. With the progress of minimally invasive diagnostics and treatment techniques, these medical tools must not only be thin or small, but must also be capable of performing several functions. To meet these demands, microfabrication techniques such as MEMS (micro electromechanical systems) technology and related microfabrication technologies are effective. A MEMS device can incorporate several functions, such as sensor, actuator, and microelectronics. We have developed several microdevices for minimally invasive diagnostics and therapies using MEMS and related microfabrication technologies. An ultraminiature fiber-optic pressure sensor and intraluminal MRI (magnetic resonance imaging) probe have been developed for use in the human body. Active bending catheters and endoscopes using micro actuators have been developed for steering their tips without using traction of wires from outside the body. Small wearable healthcare devices which are attached on the surface of the human body have also been developed using the microfabrication technologies. A wearable blood pressure monitoring system using ultrasound and a microperfusion system using a metal needle with micro flow channel for measurement of subepidermal biological substances have been developed.

Keywords Endoscope • Catheter • Medical device • Healthcare device • MEMS technology

Y. Haga (✉) • T. Kobayashi • N. Tsuruoka
Graduate School of Biomedical Engineering, Tohoku University, Sendai, Japan
e-mail: haga@bme.tohoku.ac.jp

T. Matsunaga
Micro System Integration Center, Tohoku University, Sendai, Japan

1   Introduction

Micro electromechanical systems (MEMS) is one of microsystems technology for building micromechanical structures on wafers using photofabrication techniques, for example, photolithography, etching, patterning, and bonding. One of advantages of the MEMS process is batch fabrication which enables disposable use of the device. The advantage is suitability for minimally invasive therapy because almost the tools for use in the human body are disposable.

Minimally invasive examinations and therapies with endoscopes and catheters are already widely performed, and enable effective treatment without large incision and damage to the patients. There are mainly three kinds of tools to perform minimally invasive therapy:

- Catheters, flexible thin tubes which are inserted in the blood vessels or digestive tracts for intraluminal examinations and surgery.
- Rigid endoscopes, which are inserted through small incisions of the skin.
- Flexible endoscopes, which are inserted into natural openings of the human body.

In the case of the catheter, precise position control and effective imaging method are required for precise intravascular surgeries, for example, RF catheter ablation and embolization of cerebral aneurysms. Precise endoscopic surgery techniques are required for advanced endoscopic treatment procedure, for example, ESD (Endoscopic Submucosal Dissection) as treatment of cancer.

To meet the requirements, microsystem technologies are effective. Using these technologies, several small medical devices and systems with several functions for use in the human body can be realized.

Other applications of thin and small devices with several functions are healthcare applications. Small wearable healthcare devices which are attached on the surface of the human body have also been developed using the microfabrication technologies.

2   Microdevices Developed for Minimally Invasive Diagnostics and Treatment

Using these microfabrication technologies, we have developed several small and thin medical devices and systems with several functions for use in the human body.

2.1   *Ultraminiature Fiber-Optic Pressure Sensor*

Small pressure sensors are effective for accurate local pressure measurement in small blood vessels of small children or animals, and in stenosed diseased areas. Fiber-optic pressure sensors have the advantages of high potential of miniaturization. Figure 15.1 shows a sensor system [1]. White light passes through the fiber
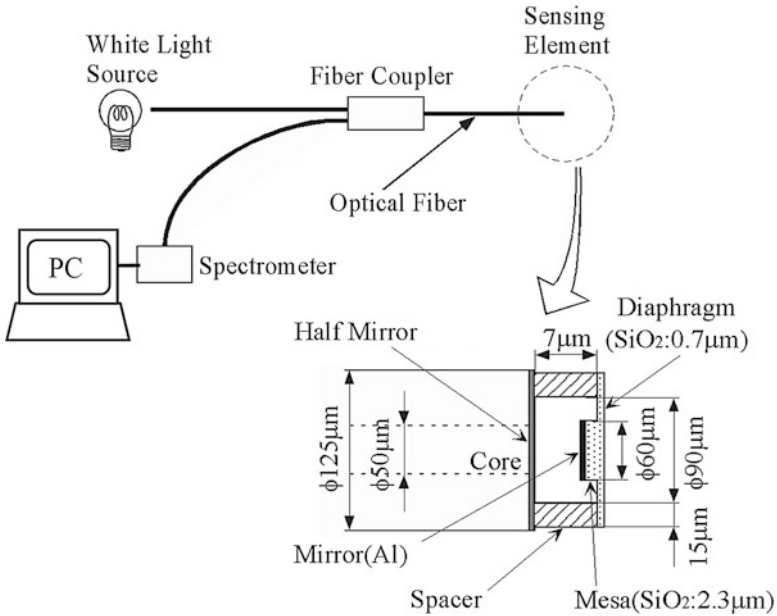
Fig. 15.1 Sensor system of ultraminiature fiber-optic pressure sensor [1]

Fig. 15.2 Fabricated sensor head [1]
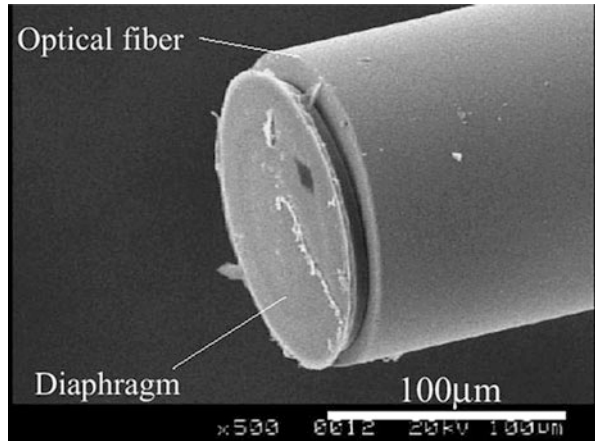
to the sensor and light reflected by the sensor element passes back through the same optical fiber and is finally detected by a spectrometer. As a sensor element, a small thin SiO_2 diaphragm is fabricated and attached at the end of the optical fiber as shown in Fig. 15.2. The thickness of the SiO_2 diaphragm is 700 nm. A total reflection mirror is fabricated on the backside of the diaphragm and a half mirror is also formed at the end of the optical fiber. The diaphragm is deformed by applied blood pressure and white light is interferometrically modulated according to the

gap change between the two mirrors. As the sensor head is small enough, several sensor elements are batch fabricated at one fabrication process on a silicon wafer and each sensor element is attached to the fiber end in a glass capillary. For use in blood vessels, the fiber-optic pressure sensor is covered with a metal coil and a biocompatible polymer tube. Experiments on animals have been carried out and blood pressure in porcine artery was successfully measured.

2.2 Endoscopic Hood with Hydraulically Variable Tip Diameter for Endoscopic Submucosal Dissection

Endoscopic submucosal dissection (ESD) is performed widely for treatment of early stage cancer in the digestive tract. Early cancer is accurately dissected en bloc by ESD regardless of its size, but it is a technically demanding procedure with a risk of complication. Hood with hydraulically variable tip diameter has been developed for lifting up the incised mucosa and attaining large field of endoscopic view to reduce the operation time and the risk of perforation [2].

The tip parts and the base part are made of acrylic resin using a 3D printer. The hood has its tip divided into two or three parts, and the tip parts and the base parts are joined by polymer sheets. A thin silicone tube which acts as a hydraulic actuator is embedded inside the hood. When water is injected inside the silicone tube using a medical syringe, the polymer sheets deform by the inflation of silicone tube, and the tip parts are pulled open, hence the tip diameter of the hood changes.

The hood with variable tip diameter is mounted at the tip of the endoscope (Fig. 15.3) and inserted into the incision when its tip diameter is small. Then, the
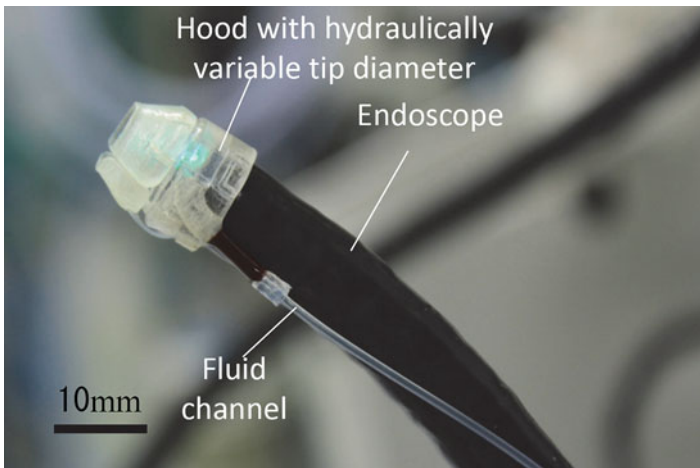
Fig. 15.3 Fabricated endoscopic hood with hydraulically variable tip diameter mounted at the tip of the endoscope [2]
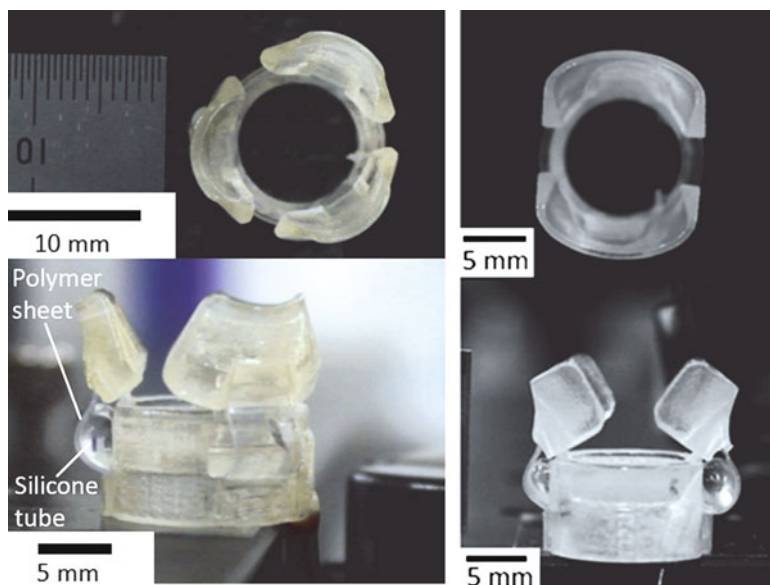
Fig. 15.4 Actuated endoscopic hood with hydraulically variable tip diameter, three tips type (*left*) and two tips type (*right*) [2]

lesion is lifted by expanding the tip diameter, widening the visible incision area as shown in Fig. 15.4.

2.3   Active Catheter/Active Endoscope Using Shape Memory Alloy Actuators

For controllable steering of the catheter or the endoscope, TiNi alloy shape memory alloy (SMA) micro actuators can be used. The SMA is heated and actuated by supplying electrical current to the SMA wire when electrical resistance of the SMA wire is relatively high if the wire diameter is small. Then the SMAs can be easily actuated as small artificial muscles. By changing the deformation shape of the SMA and configuration of the mechanism, several motions can be realized [3]. Not only bending motion but also torsional and extensional motions can be realized as shown in Fig. 15.5.

Multidirectional bending mechanism using SMA wires has been developed and utilized for an active bending electric endoscope [4]. Three SMA wires are positioned with a spacing of 120° around the inner tube. When one or two SMA wires are actuated, they contract and bend the mechanism in a certain direction. To prevent deterioration of bending angle caused by inactive SMA wires, the inactive SMA wires shift inward without strain as shown in Fig. 15.6.
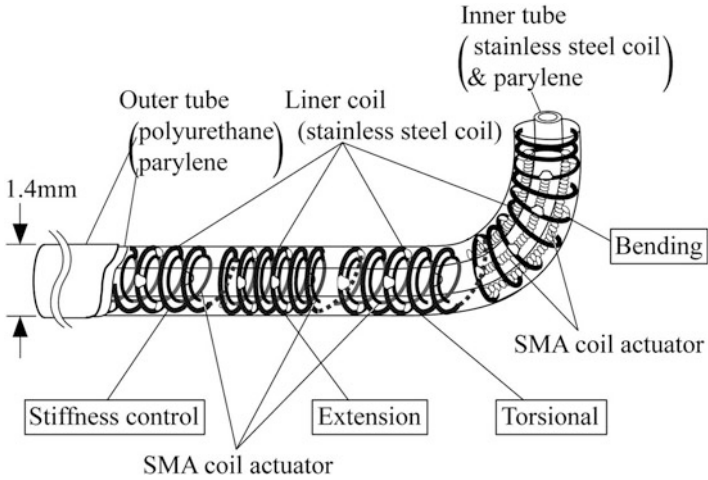
Fig. 15.5 Conceptual structure of active mechanisms using SMA [3]

Fig. 15.6 Multidirectional bending mechanism using SMA wires [4]
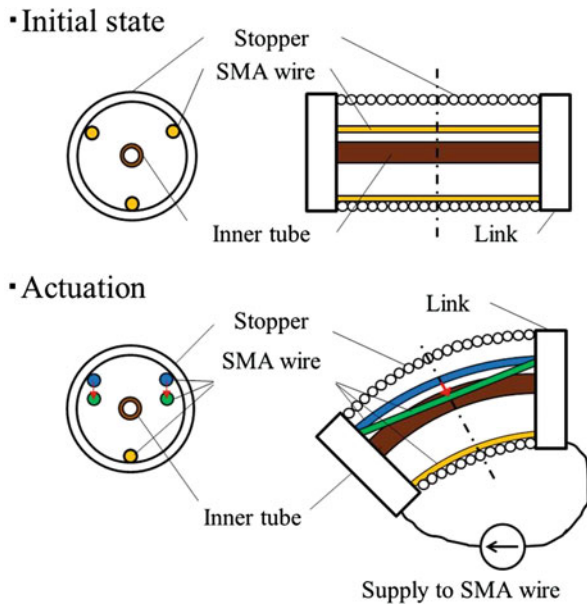
An active bending electric endoscope using the multidirectional bending mechanism has been fabricated as shown in Fig. 15.7 [4]. The endoscope consists of an imager component, a bending component, and a shaft component. The imager component has a CMOS imager, and three LEDs. The bending part consists of three SMA wires, a restraining coil made of stainless steel wire with a square cross section, and four links for adequate positioning of the SMA wires. The bending angle of the active bending mechanism was 50° (radius of curvature 40 mm) at
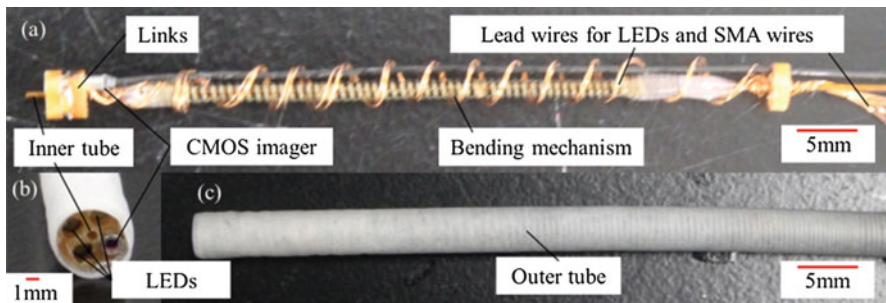
Fig. 15.7 Fabricated active bending electric endoscope using SMA wires [4]
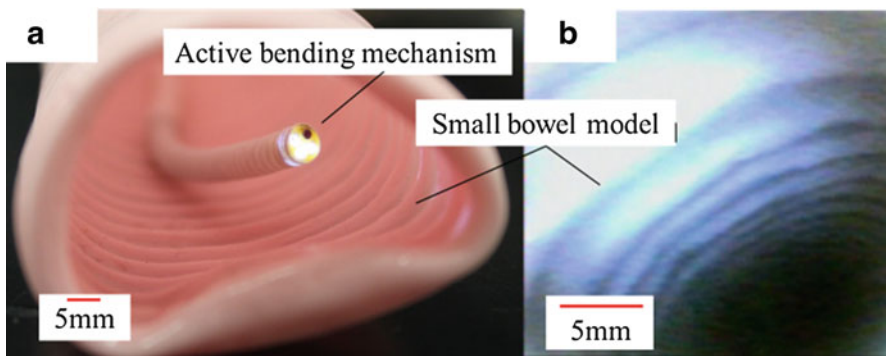
Fig. 15.8 Fabricated endoscope in bowel model (**a**) and acquired image of the bowel model using equipped CMOS imager (**b**) [4]

300 mA when one SMA wire was actuated. Fabricated active bending electric endoscope in small bowel model and acquired image of inside the bowel model using equipped CMOS imager are shown in Fig. 15.8.

2.4 High-Resolution Intraluminal Magnetic Resonance Imaging Probe

Intraluminal MRI (magnetic resonance imaging) probe holds promise to achieve high-resolution image of small pathological lesion such as the vessel plaque and pancreas cancer comparing to conventional MRI scanner.

By placing a receive coil in the human body, NMR (nuclear magnetic resonance) signals from the tissue surrounding the receive coil are detected sensitively.

For making the receive coils, we have developed a MEMS process applicable to cylindrical substrates. A point exposure system has been developed using maskless photolithography on cylindrical substrates; a sputtering system with a rotational
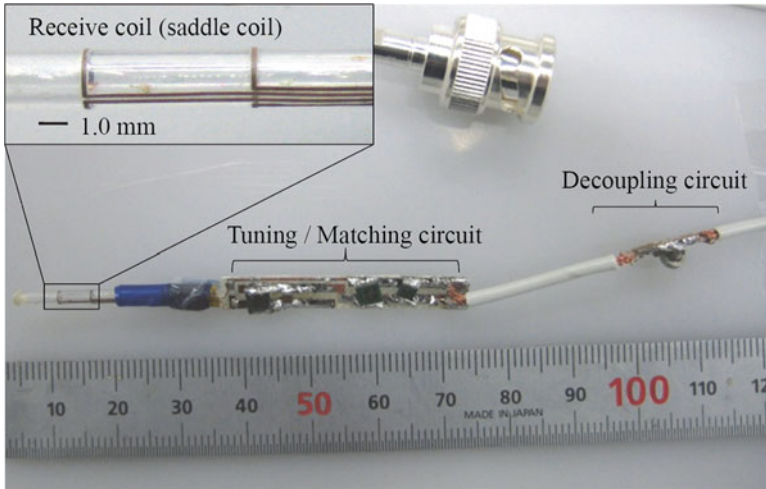
Fig. 15.9 MRI probe (single layer saddle coil) [5]

mechanism and a photoresist spray coating system have been fabricated and utilized. Arbitrary thick photoresist pattern can be fabricated on the cylindrical substrates using the equipments. Thick metal pattern is electroplated using the thick photoresist pattern as a mold as shown in Fig. 15.9 [5].

Fabricated small receive coil probe is inserted into blood vessel. Acquired image of extracted porcine subclavian artery is shown in Fig. 15.10. Multilayer coil pattern is effective for improvement of sensitivity by increasing inductance of the coil. Multilayer coil pattern fabricated by repeating the process mentioned above is shown in Fig. 15.11.

3 Healthcare Devices Using Microsystem Technology

3.1 Wearable Blood Pressure Sensor Sheet Using Ultrasonic Measurement of Vessel Diameter

For wearable blood pressure monitoring without cuff, ultrasonic transducer array which measures change of blood vessel diameter has been developed as shown in Fig. 15.12 [6]. The transducer array is placed on the skin of the wrist above radial artery.

When stiffness of the vessel wall is constant, diameter of the blood vessel changes according to internal blood pressure in the blood vessel. The diameter change can be measured using delay time of reflected ultrasonic pulse echo signal. After initial calibration of maximum and minimum diameter of the blood vessel
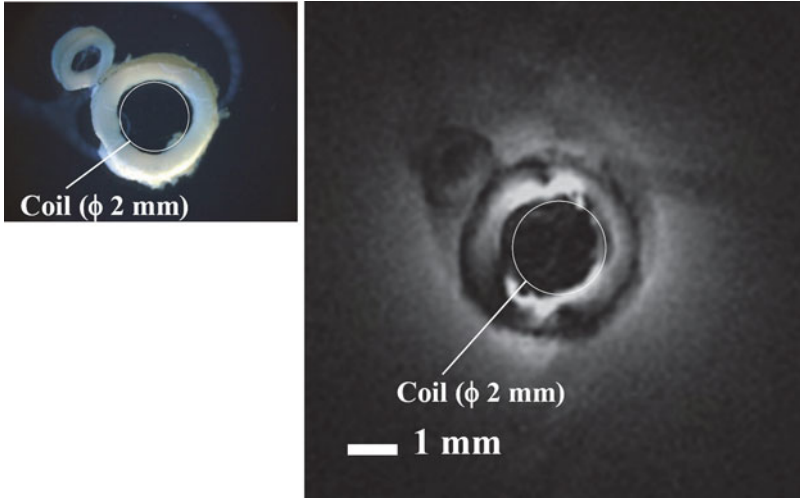
Fig. 15.10 Extracted porcine subclavian artery (*left*) and acquired MRI image (*right*) [5]
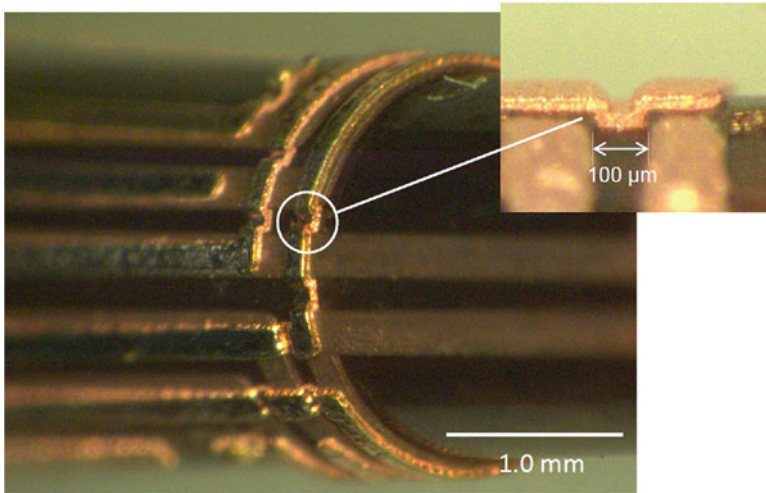
Fig. 15.11 Multilayer coil

using systolic blood pressure and diastolic blood pressure measured by conventional blood pressure cuff, blood pressure can be estimated using diameter change.

Figure 15.11 shows structure and fabricated ultrasonic transducer array. Array of the transducer is useful for alignment of ultrasonic transducer and radial artery. Using the device, displacements of the posterior and anterior vessel walls were measured, respectively. Diameter of blood vessel is calculated by subtraction of two wall displacements.
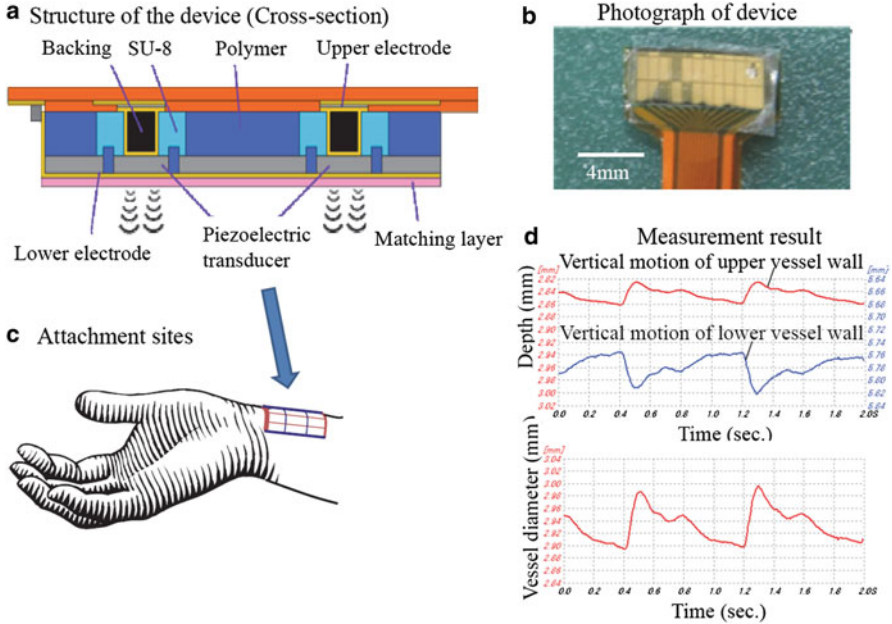
Fig. 15.12 Wearable blood pressure sensor sheet using ultrasonic measurement of vessel diameter [6]
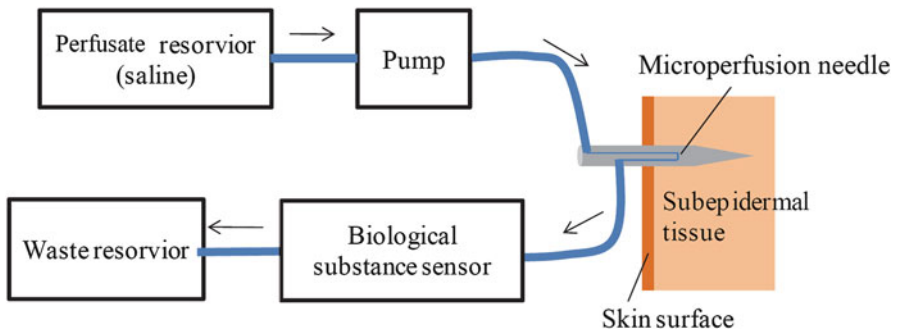
Fig. 15.13 Microperfusion for measurement of biological substances in subepidermal tissue [7]

3.2 Microperfusion Needle for Measurement of Biological Substances in Subepidermal Tissue

Concentrations of biological substances are useful as indicators of physiological and pathological states. In order to monitor biological substances, thin metal needle which has fluid channel with perforated membrane has been developed as microperfusion needle as shown in Fig. 15.13. Biological substances from
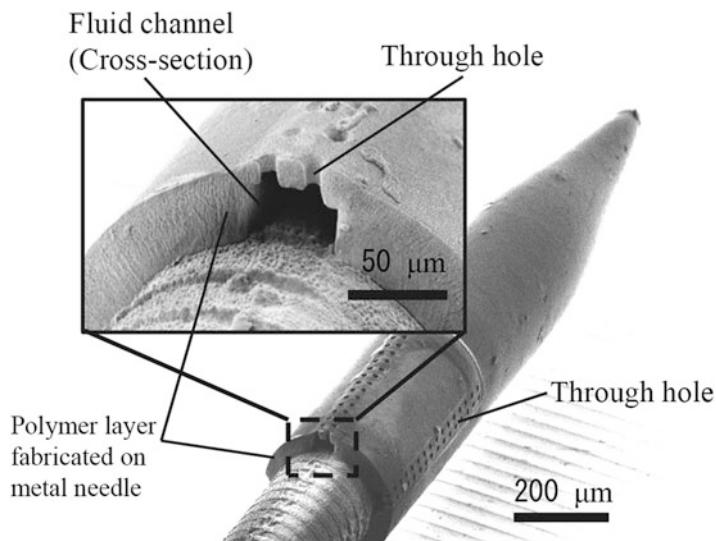
Fig. 15.14 Fabricated microperfusion needle (fluid channel is fabricated on surface of acupuncture metal needle) [7]

subepidermal tissue are extracted by perfusion of saline. Concentrations of some low molecular weight substances in blood, for example, blood glucose and blood lactate, are well correlated to concentrations in subcutaneous tissue.

For microperfusion, fluid channel with perforated membrane is fabricated on the surface of the 200 μm diameter metal needle as shown in Fig. 15.14 [7] using the non-planer photofabrication process mentioned in Sect. 2.4. Since this device uses thin metal acupuncture needle as base substrate, the needle has enough rigidity for insertion through the skin and can be inserted in the skin without pain.

The device can be used for several applications, for example, adequate exercise load level indicator using lactate monitoring, alcohol sensor, and glucose sensor for diabetes management. The efficacy of the fabricated needle was examined by measuring lactate and glucose concentrations of mice. Concentrations in perfused saline were well correlated to concentrations in blood.

4   Conclusion

Various biomedical microsystems for minimally invasive diagnosis and treatment and for healthcare applications have been mentioned. Minimally invasive diagnosis and treatment and healthcare are promising applications of microsystem technologies, not only because the technologies make conventional procedures easier and more precise, but also because there is a demand for technology which can enable improved diagnosis and treatment.

References

1. Totsu K, Haga Y, Esashi M (2005) Ultra-miniature fiber-optic pressure sensor using white light interferometry. J Micromech Microeng 15(1):71–75
2. Matsuo M, Iijima K, Matsunaga T, Haga Y (2015) Development of hood with hydraulically variable tip diameter for endoscopic submucosal dissection. Sensors Actuators A 232:267–275
3. Haga Y, Mineta T, Totsu K, Makishi W, Esashi M (2001) Development of active catheter, active guide wire and micro sensor systems. Interv Neuroradiol 7:125–130
4. Kobayashi T, Matsunaga T, Haga Y (2015) Active bending electric endoscope using shape memory alloy wires. In: Bleuer H et al (eds) New trends in medical and service robotics. Springer, Dordrecht, pp 131–139
5. Goto S, Matsunaga T, Matsuoka Y, Kuroda K, Esashi M, Haga Y (2008) Development of intravascular MRI probe applicable to catheter mounting. IEEJ Trans Sensors Micromach 128(10):389–395 (in Japanese)
6. Suzuki D, Nishiyachi K, Fujita K, Maehira K, Oohashi Y, Fuwa K, Matsunaga T, Hasegawa H, Kanai H, Haga Y (2012) Development of wearable blood pressure sensor using ultrasound measurement. In: Proceedings of annual conference of Japanese society for medical and biological engineering (in Japanese)
7. Tsuruoka N, Ishii K, Matsunaga T, Nagatomi R, Haga Y (2016) Lactate and glucose measurement in subepidermal tissue using minimally invasive microperfusion needle. Biomed Microdevices 18:19

Chapter 16
Flexible Electronic Devices for Biomedical Applications

Pooria Mostafalu, Amir Sanati Nezhad, Mehdi Nikkhah, and Mohsen Akbari

Abstract Recent advances in material sciences and microfabrication technologies have enabled creating flexible electronic systems that are able to integrate with soft tissues with curvilinear and dynamic surfaces. Remarkable features of these flexible devices have opened an array of opportunities in controlling disease conditions, improving surgical procedures, and continuous health monitoring. The overarching goal of this chapter is to provide an overview of current advances in the field of flexible electronics with emphasize on biomedical applications. We will primarily discuss the fabrication strategies and materials for the development of physical-, chemical-, and biosensors. In the second part, the emerging applications of flexible electronics in wound healing, wearable electronics, implantable devices, and surgical tools as well as point-of-care diagnostic devices, will be explored.

Keywords BioMEMS • Flexible electronics • Biomedical engineering • Biosensors • Wound healing • Implantable devices • Surgical tools • Wearable electronics • Physical sensors • Chemical sensors • Point-of-care diagnostics

P. Mostafalu
Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02139, USA

A. Sanati Nezhad
Department of Mechanical and Manufacturing Engineering, BioMEMS and Bioinspired Microfluidic Laboratory, Schulich School of Engineering, University of Calgary, EEEL 455B, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada

M. Nikkhah
School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287, USA

M. Akbari (✉)
Department of Mechanical Engineering, Laboratory for Innovation in MicroEngineering (LiME), University of Victoria, 3800 Finnerty Rd., Victoria, BC V8P 2C5, Canada

Center for Biomedical Research, University of Victoria, 3800 Finnerty Rd., Victoria, BC, Canada, V8P 2C5, Canada
e-mail: makbari@uvic.ca

1 Introduction

Flexible electronics is referred to the technology of integrating electronic circuits on stretchable substrates. The introduction of this technology goes back to 1960s when the flexible solar cell arrays were fabricated by forming thin layers ($\sim$100 μm) of crystal silicon wafer cells on a plastic substrate [1]. Flexible electronic devices offer several advantages over their rigid counterparts: they are lighter, portable, and could be manufactured using less expensive methods. Additionally, these devices are foldable and deformable, which make them excellent for applications that require smaller footprints or possess surfaces with curvilinear and irregular morphology. For instance, flexible solar cells are now being widely used to power satellites as they are lightweight and can be rolled up during the launch time and expanded after the satellite is placed in its orbit [2, 3]. Because of these alluring properties, this technology has been rapidly grown and found extensive applications in automotive industry [4], energy harvesting [5, 6], paper-like displays [7–9], robotics [10, 11], and medical devices [12–14].

Flexible electronics can be integrated with biological tissues by conforming to their soft and curvilinear shapes [13–15]. These systems can be bent, twisted, and wrapped around irregularly shaped objects without compromising their functionalities [13]. Due to these unique properties, innovative flexible electronic systems have been developed to measure electrophysiological signals [14, 16–19] and deliver therapeutic agents into or onto the human body [20, 21]. Additionally, recent advances in developing novel synthetic and naturally derived biomaterials enabled tailoring the substrate mechanical properties and controlling the rate of adsorption of different components of electronic systems in a programmable manner for implantable applications.

This chapter provides an overview of the recent advances in flexible electronic systems from the perspectives of their applications in biomedical applications. The first part of the chapter discusses flexible biomaterials including biocompatible substrates and conductive materials that have been utilized for this technology. We will explore different fabrication methods for creating flexible electrodes and will discuss their applications in the development of physical-, chemical-, and biosensors. The second part of this chapter will be dedicated to the emerging applications of the flexible electronics in wound healing, implantable devices, wearable electronics, and point-of-care diagnostic devices for early detection of human diseases. The chapter finally presents a perspective on the technology of flexible electronics and provides insights for the development of novel systems for emerging applications in biomedical engineering.

2   Flexible Electronics

2.1   Fabrication Strategies and Materials

Modern electronic devices are composed of a circuitry system made of semiconductors and a substrate that may be encapsulated in epoxy-based materials to protect delicate components. In flexible electronic systems, all components should comply with arbitrary deformations such as bending, stretching, twisting, and compression without losing their performance. One of the major strategies for fabricating flexible electronics is to manufacture complete circuits on a carrier substrate using conventional methods including lithography-based patterning techniques or chemical etching procedures, and further transferring the patterns onto a flexible substrate [22, 23]. In a study by Choi et al., silicon nanomembranes were fabricated from silicon-on-insulator wafers using reactive ion etching and the structures were subsequently transferred and bonded to a pre-strained polydimethylsiloxane (PDMS) substrate [22] (Fig. 16.1a). This construct exhibited biaxial and reversible stretching and compression with a linear response to applied forces. To improve the stretchability of the structures, they were formed into a noncoplanar mesh architecture to move freely out of the plane [23] (Fig. 16.1b, c). With such configuration, strains up to 100 % were achieved. Other structures such as noncoplanar serpentine [23] and coiled spring layouts [24] have also been used to achieve higher stretchability. Although this process enables the fabrication of high-performance devices with sophisticated structures on flexible substrates, it is expensive and can only cover small surface areas. Moreover, the manufacturing process is challenging due to small dimensions and mechanical fragility of the features.

An alternative approach for the development of flexible electronics is to fabricate the circuits directly on a flexible substrate. This approach requires conductive materials that can be grown on a foreign substrate and stretched without losing their functionalities. Existing techniques for direct patterning of conductive polymers include low-temperature assembly [25–27], solution processing and direct deposition [28, 29], contact printing using PDMS stamps [17, 30, 31], screen printing [32, 33], micro- and nanomoulding [12, 34, 35], and embedded 3D printing [36]. Typical organic/inorganic materials used in these techniques include eutectic gallium indium (eGaIn) [37, 38], carbon-based materials [39, 40], nanowires or nanotubes [39, 41], silver [42], and ionic fluids such as hydrogels and ionogels [43, 44]. A major advantage of these methods is their scalability that allows manufacturing of large-area electronic surfaces with high density. For instance, a 3D printing approach was utilized to directly embed a conductive material within a flexible substrate [36]. This method was used to create strain sensors for human-motion detection (Fig. 16.2a–c) [36].

Roll-to-roll fabrication of large-scale flexible electronic systems was demonstrated by direct deposition of monolayers of graphene films onto flexible copper substrates [29] and conductive polymers ITO:PET substrate [45] (Fig. 16.2d–f). However, any of these materials has its own limitations. For example, the fabrication
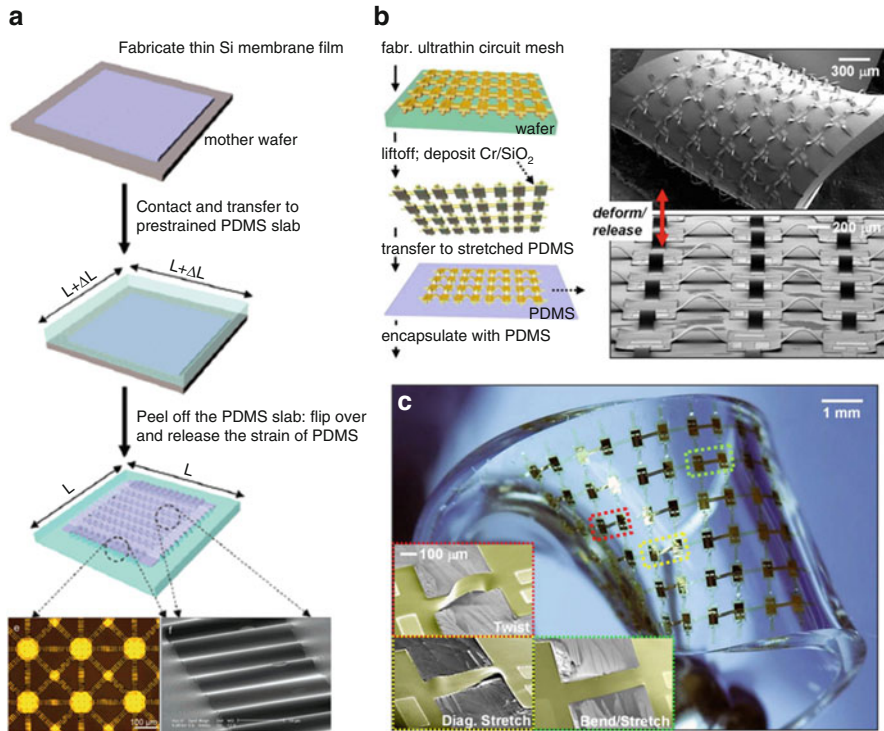
Fig. 16.1 The stretchable features transferred and bonded to elastic substrates. (**a**) Fabrication and transfer of wavy features on PDMS. Reprinted with permission from [22]. Copyright 2007 by American Society of Chemistry. (**b**) Fabrication process of noncoplanar stretchable electronics. The circuit mesh was fabricated on a silicon wafer and then transported to a pre-strained PDMS substrate and encapsulated within PDMS. (**c**) Flexible device underwent various stretching modes including diagonal stretch, twisting, and bending. Reprinted with permission from [23] Copyright 2008 by Proceedings of National Academy of Sciences

routs of using liquid metals are complex as they possess high surface tension and complex surface phenomenon [46]. On the other hand, ionic fluids and conductive inks can be patterned easier, but they have non-uniform electrical resistivity because of particle network disruption at high strains [36] or changes in water content as a result of evaporation or hygroscopic effects [47].

Flexible electronics with the ability of conforming to curvilinear surfaces require patterning of electronic components on stretchable substrates [48]. Metallic free-standing microstrips have been successfully utilized for brain mapping [14]. However, having a flexible substrate improves the mechanical stability of the device and enables fabrication of advanced drug delivery systems and smart wound dressings by encapsulating drug carriers in the substrate. Paper and PDMS are probably the most used substrates in flexible electronic devices [13, 23, 49, 50]. These materials are biocompatible, low cost, and permeable to air. In particular, the nanoporous structure of paper makes it an excellent candidate for rapid diagnostic
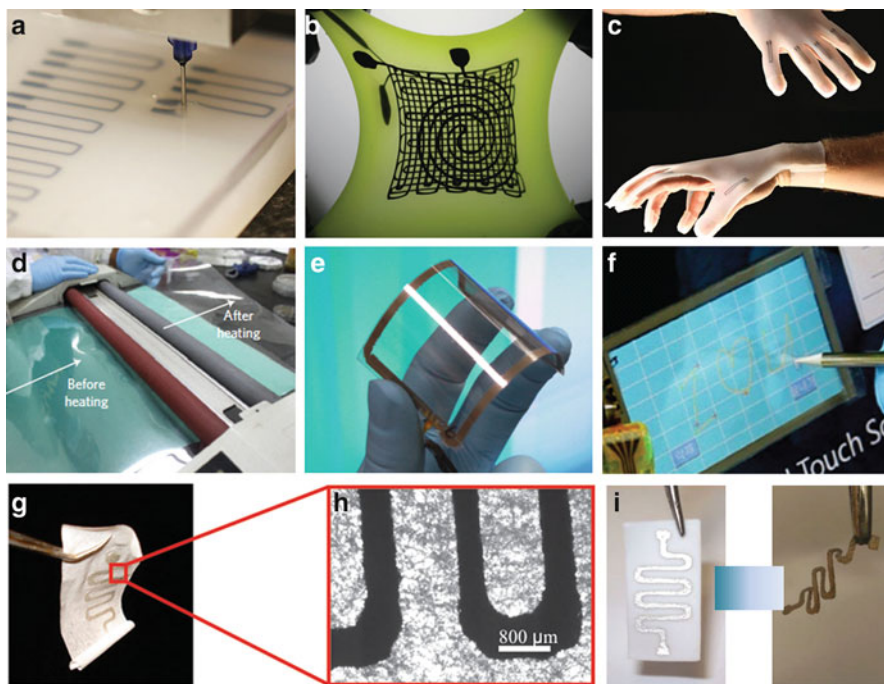
Fig. 16.2 Manufacturing of flexible electronics. (**a**) Embedded 3D printing of conductive ink into an elastic substrate. (**b**) Image of a stretched multilayer strain and pressure sensors fabricated by embedded 3D printing technique. (**c**) A glove with strain sensors made by embedded 3D printing. Reprinted with permission from [36]. Copyright 2014 by WILEY-VCH Verlag GmbH & Co. (**d**) Roll-to-roll fabrication of a large-scale flexible electronic device with (**e**) excellent flexibility. (**f**) A touch screen panel resulted from the graphene-based stretchable electronic. Reprinted with permission from [29]. Copyright 2010 by Macmillan Publishers Limited. (**g**) A patterned electrode on an electrospun PGS-PCL mat. (**h**) Scanning electron microscopy image of the electrode on the mat. (**i**) Flexible electrode before (*left*) and after (*right*) degradation of the PGS-PCL substrate in 500 mM NaOH. Reprinted with permission from [32]. Copyright 2014 by WILEY-VCH Verlag GmbH & Co

tests. However, paper is not elastic and loses its mechanical integrity in aqueous environments. PDMS is a soft elastomer that has been extensively used in food and cosmetic industry. PDMS can be processed easily to create stamps for transferring conductive materials on different substrates [51]. Due to the existence of –OH groups on the PDMS surface, strong covalent bonds (Si–O–Si) between the pattern on the top and the substrate can be formed. Other materials such as silk [52], polyester [15], poly(imide) [53], poly(4-vinylpyridine) [54], paper, poly(lactic-*co*-glycolic) acid (PLGA) [55], and poly(caprolactone)-poly(glycerol sebacate) (PGS-PCL) [32] have also been used as substrate materials. A notable example is an electrospun PGS-PCL mat used to fabricate bio-resorbable temperature and strain sensors and heaters by screen-printing of silver ink and iron oxide (Fig. 16.2g–i) [32]. These fibrous substrates offer several advantages such as high

biodegradability, considerable suturability, high in-plan thermal conductivity, and high permeability to air and body fluids [32, 56, 57]. Similar to paper substrates, electrically conductive materials can be patterned on these substrates using screen printing, inkjet printing, and contact printing.

Overall, elastic electronics that are able to form a conformal contact with non-flat surfaces require an elastic substrate and stretchable conductive materials. The fabrication process should maintain the contact between the conductive layer and the substrate without losing its performance. For biomedical applications, the material should also be biocompatible and does not induce allergic reactions. Implantable electronics should degrade into non-toxic compounds in a controlled manner. Moreover, these materials should be compatible with the existing fabrication processes and enable the fabrication of large-scale devices.

2.2 Physical Sensors

Physical parameters such as temperature, stress, strain, and pressure are important indicators of different diseased conditions in various organs and tissues [58–60].

2.2.1 Temperature Sensors

Temperature is an important marker for monitoring the body's health as its variation can be an indicator of inflammation or bacterial infection [58, 59]. Moreover, precise measurement of significant changes in temperature is central for improving the accuracy of other chemical and physical sensors [61]. For the majority of temperature sensors, the principle operation is based on the variation of the resistance as a function of temperature change. Resistive metallic sensors are the most common types of sensors, for this purpose [61]. Platinum is the best material among readily available metals that provides a linear temperature-resistance coefficient and is compatible with conventional microelectromechanical systems (MEMS) technology [15, 62–64]. Other materials such as graphite, silver, and carbon nanotubes (CNTs) have also been widely used for the fabrication of flexible temperature sensors [65–68].

2.2.2 Strain Sensors

Strain sensors are used for measuring the amount of displacement in a moving object. The basic of strain sensors relies on the piezoresistive properties of a conductor, where the electrical conductivity varies with applied strains mostly in a linear fashion [15, 39]. A mechanotransduction system inspired from kidney cells relies on the mechanical interlocking of metal coated nanofibers [12]. Such nanofibers can be made from high-aspect ratio Pt-coater polymeric nanofibers grown on thin layers of PDMS substrates (Fig. 16.3a). This configuration can be used to measure strains as small as human heart beats and as large as the impact of a water droplet on a super-hydrophobic surface with reproducibility of 10,000 cycles
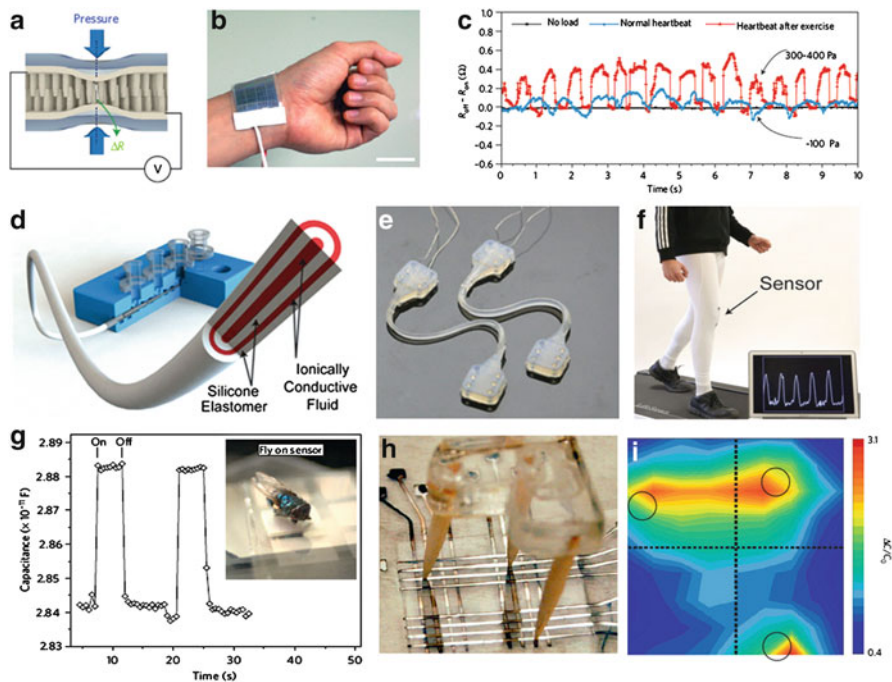
Fig. 16.3 Flexible physical sensors. (**a**) Schematic of the strain sensor made from high-aspect ratio Pt-coater polymeric nanofibers grown on thin layers of PDMS, which was used for (**b**, **c**) human heart beat measurement. Reprinted with permission from [12]. Copyright 2012 by Macmillan Publishers Limited. (**d**) Schematic of a multicore-shell capacitive strain sensor for human motion measurement. (**e**) Multicore-shell strain sensors fabricated in two different lengths. (**f**) A stain sensor integrated with a fabric across the knee for human motion measurement. Reprinted with permission from [69]. Copyright 2015 by WILEY-VCH Verlag GmbH & Co. Ultrasensitive flexible pressure sensors capable of measuring ultralight weights of a (**g**) fly and (**h**, **i**) miniaturized tripod. Reprinted with permission from [35]. Copyright 2010 by Macmillan Publishers Limited

(Fig. 16.3b, c) [12]. Embedded 3D printing of a viscoelastic conductive material into an elastomer using a nozzle has been also used to create flexible strain sensors capable of measuring strains up to 450 % [36]. Capacitive sensors are another class of strain sensors, which have been recently used to measure strains up to 700 % [69]. These sensors are composed of a dielectric layer sandwiched between two conductive layers. The strain signal is determined by measuring the alteration in the thickness of the layers as a function of applied deformation [43, 69, 70]. In particular, strain sensors that are made from fibre materials can be integrated into smart fabrics and further used for the measurement of human motion (Fig. 16.3d–f).

2.2.3 Pressure Sensors

Pressure measurement is essential for applications in soft robotics, smart artificial hands, and electronic skin (e-skin) that imitates the functions of human skin [8,

35]. To sense the applied pressure on a large surface, fabrication of an array of flexible pressure sensors capable of mimicking the tactile sensing of natural skin is required. Organic field-effect transistors (FETs) integrated with rubber pressure sensors are the most cost-effective methods for fabrication of large-scale flexible pressure sensors [10, 11]. Organic transistors are inherently flexible and offer several advantages including the ability to be manufactured on plastic films at ambient temperatures using inexpensive methods [53, 71, 72]. Higher sensitivities can be achieved by using a compressible rubber in the dielectric layer (Fig. 16.3g–i) [35].

Integration of the sensors with other electronic components such as power supplies and communication elements is central for long-term measurement of the physical markers [15]. An important factor in designing such systems is to ensure long-live and robust electrical contacts that do not cause discomfort and irritation for the skin. In a notable study, a multifunctional sensing system composed of temperature, strain, and electrophysiological sensors, light-emitting diodes, transistors, resistors, wireless power coils, along with a high frequency communications [15]. This sophisticated system was integrated on a 30 µm gas permeable polyester sheet and was able to undergo different stretching forces without losing its functionalities. Such advanced technology can be further used to measure the electrical signals generated by heart, brain, and skeletal muscles for gathering information about the human body health in long term.

2.3 Chemical and Biological Sensors

Flexible chemical- and biological sensors designed for the detection of biological parameters such as pH, biomarkers, enzymes, chemokines, toxic or explosive agents, or environmental pollutants. These sensors have been used for a wide range of applications including healthcare, military, sport, and environmental monitoring [73]. Development of miniaturized and minimally invasive biosensors for the detection of biomarkers and cytokines continues to be an important application of flexible biosensors [73]. Among a variety of different detection systems, electrochemical detection techniques such as amperometric and voltammetric methods have received a broad attention. Various flexible materials including polyethylene terephthalate (PET), polyethylene naphthalate (PEN), Mylar, Parylene, and Kapton have been used as the substrate for electrode patterning [74]. Graphene is an exceptional candidate as an electrode for flexible biosensors due to its high deformability, conductivity, and transparency [17]. Also its high surface area helps to provide high sensitivity for a large number of markers [75]. To use graphene for biosensing applications, its surface can be functionalized with various biomolecules, e.g., antibodies and linkers through covalent, non-covalent, or ionic bonds [76]. In the most common application of graphene for biosensing, the surface is functionalized with reduced graphene oxide (rGO) [77, 78].

2.3.1 pH Sensors

pH is one of the critical parameters affecting many reactions in the body and is an indicator of the bacterial infection especially in chronic wounds [79]. The pH of digestive tract also changes in gastrointestinal diseases such as inflammatory bowel and gastroesophageal reflux diseases or infection from *Helicobacter pylori* [80, 81]. Traditional pH sensors are glass-based potentiometric devices that reveal open circuit voltage with respect to the pH of the solution based on Nernst equation [82]. These devices are rigid and bulky, and cannot be used for wearable and in vivo applications. Solid-state pH sensors contain two electrodes, a working electrode and a reference electrode. Working electrode is fabricated by deposition of metal oxide such as platinum oxide, iridium oxide, and ruthenium oxide on flexible substrates [83–86]. These metals can be deposited using physical metal deposition including sputtering, e-beam evaporation, electroplating, screen printing, and sol–gel coating [85, 86]. The thickness of the deposited layer is small and does not affect the flexibility of the device. Conductive polymers such as polyaniline (PANI) or polypyrrole (PPy) are other materials that could be used as working electrodes in pH sensors [87, 88]. Their chemical structure allows protonation and deprotonation in acidic and basic environments to generate the potential voltage between working and reference electrodes [89]. Solid-state pH sensors demonstrated a linear response within the pH range of 2.0–12.0 with a sensitivity of 58.73 ± 0.41 mV/pH and a rapid response time of 0.5–5 s depending on the solution pH [90].

2.3.2 Glucose Sensors

The ability to monitor the glucose level in vivo over a long period of time enables intensive control of blood glucose concentrations in diabetic patients. Additionally, such long-lasting implantable sensors reduce the frequency of implantation and replacement, which can lead to less discomfort for patients. A flexible thick-film biosensor was developed using Nafion/GOx/carbon enzyme-functionalized electrodes for in vivo detection of glucose (Fig. 16.4a) [91]. The mechanical deformation testing demonstrated reliable performance for such flexible enzyme/polymer electrode (Fig. 16.4b). It was also shown that the sensitivity of the sensor substantially enhanced following the strain application. A potential reason could be the increase in biocatalytic area for the reaction though the response was stabilized after a few bending cycles. However, the alteration in sensitivity under mechanical loading was not observed for other agents such as hydrogen peroxide, catechol, and potassium ferrocyanide. Another glucose sensor with flexible structure and suitable optical transparency was constructed by immobilizing graphene oxide (GO) with glutaraldehyde solution onto the Indium/Tin Oxide (ITO)-electrode [92]. The sensor was able to measure glucose ranges from 0.06 to 1.24 mmol/L with accurate reproducibility [92]. In another study, a glucose sensor was fabricated using membrane-entrapped GO on a flexible polyimide substrate and integrated with a pH sensor based on a PVC-based neutral carrier membrane [91]. By incorporation of
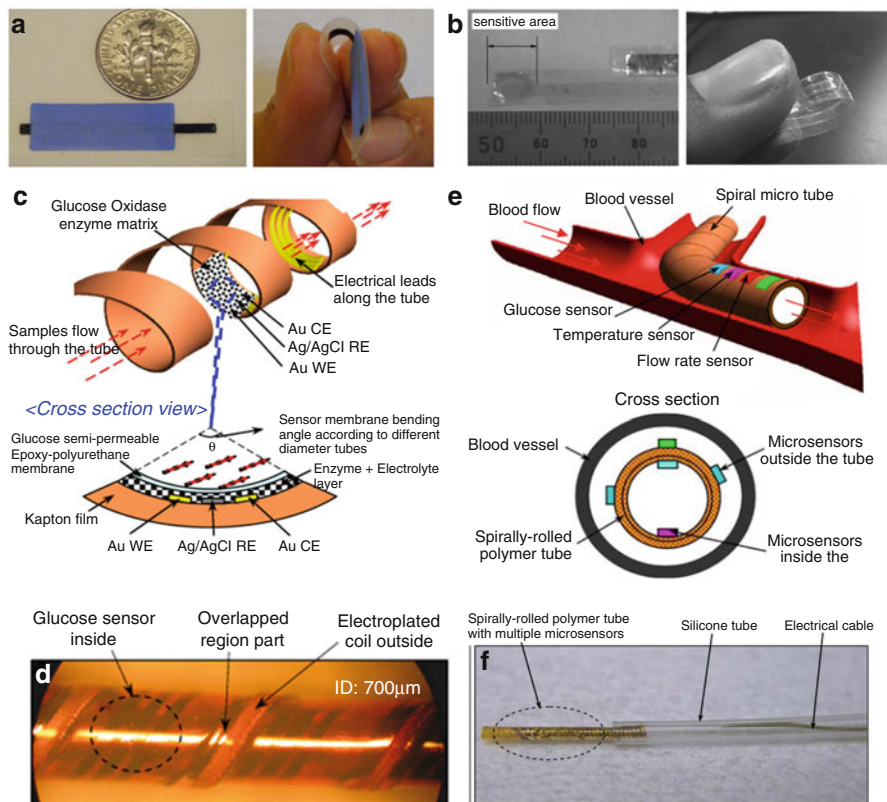
Fig. 16.4 Flexible biosensors. (**a**) Bare electrodes on a flexible Mylar substrate. (**b**) Electrodes during a 180° inward bend. Reprinted with permission from [91]. Copyright 2010 by Elsevier. (**c**) Schematic of a rolled glucose sensor and (**d**) image of the sensor. Reprinted with permission from [98]. Copyright 2007 by Elsevier. (**e**) Glucose biosensor integrated into catheter. (**f**) The fabricated glucose biosensors in rolled configuration. Reprinted with permission from [99]. Copyright 2008 by Elsevier

several pretreatment and heat curing steps as well as an electrochemical preconditioning step, the sensor performed with minimal background noise and settling time. The biosensor was designed for the short-term monitoring of glucose and pH in intensive care units particularly for neurosurgical applications. This sensor had profound performance with high specificity in detection of glucose in buffer solutions, serum, and whole blood [91].

Nanotechnology has contributed significantly to the improvement of the sensitivity and specificity of flexible biosensors. CNTs have shown promising results for the development of electrochemical biosensors due to several specific properties including strong electrocatalytic activity, minimized surface fouling, and high flexibility [3], and therefore have been used for the detection of electroactive species such as hydrogen peroxide and reduced nicotinamide adenine dinu-

cleotide (NADH). A flexible glucose sensor was fabricated using immobilization of glucose oxidase (GOx) enzyme over plasma-functionalized multiwalled carbon nanotube (MWCNT) films on gold coated PDMS substrate [93]. The GO/MWCN-T/Au/PDMS electrode exhibited highly sensitive performance within the detection limit of glucose in human body. Oxygen plasma-functionalized CNTs significantly enhanced electrocatalytic activity and electron transfer due to production of oxygen-contained groups over CNT surface, verified by X-ray photoelectron spectroscopy [18].

Temporary transfer tattoos with electrochemical-based function and compliant with the skin have been used for electrochemical analysis of physiologically relevant agents such as ascorbic and uric acids with high sensitivity [73]. The ability of these tattoos to withstand extensive mechanical loading and chemical degradation in the skin has provided reliable transdermal wearable biosensors. However, there are still challenges in the integration of electrodes onto both textile substrates and the adhesion to the epidermis. Also further improvements are required to integrate power sources to the biosensors to enable communication and data transfer via wireless infrastructure, to assess interaction of liquids at the printed electrode-epidermis interface, and to improve absorption of moisture on the surface [94].

A more advanced model of flexible biosensors is called "flexible rolled biosensors", suitable for integration into catheters for minimally invasive amperometric detection of biomarkers [95–97]. A stretchable biosensor patterned on a polyimide substrate was developed for rapid and sensitive detection of dynamic changes in glucose, norepinephrine, and dopamine [95]. To integrate such biosensors into medical catheters, textile-based thick-film sensors with highly porous substrates were effectively utilized to survive large deformations [96]. Such textile-based biosensor was further developed to achieve a flexible and low-cost solid-state sensor [97]. Since the flexible fabrics had hydrophobic properties, they prevented hysteresis behaviour and therefore enabled repeatable electrochemical sensing [97].

In a notable study, a novel glucose sensor was demonstrated by patterning electrodes on Kapton material, followed by spirally rolling it into polymer micro tube [98]. The comparison between the spirally rolled glucose biosensor with its planar counterpart showed that the rolled sensor provided reliable sensitivities under various working temperatures (Fig. 16.4c, d) [98]. This biosensor can be effectively used for in vivo monitoring of glucose using microcatheters. In a more complex system, bio-chemical and physical sensors were integrated into a flexible polymer tube mounted within a catheter for measuring physiological (temperature, pressure, and flow rate) and biochemical data (glucose concentration) (Fig. 16.4e, f) [99]. The sensors were assembled on the inside and outside wall of the polymer without the need to external wiring. The fabricated sensors demonstrated acceptable performance in planar and spiral configurations as one step forward to develop patient-specific "smart" catheters.

2.3.3 Other Biosensors

Flexible biosensors are crucial for the fast, simple, reliable, and in-field detection of biological and toxic agents on the battlefield for security applications [100–102]. In particular, electrochemical biosensors have offered a unique opportunity for field screening to identify various explosives or toxic agents [101]. Nitroaromatic explosives (e.g., 2,4,6-trinitrotoluene, TNT) with redox activity have made them ideal for electrochemical-based detection [100]. Various materials such as cotton, polyester, and Gore-Tex with different hydrophobicity have been used as flexible substrates for electrochemical detection of such agents [73]. Among these materials, fabrics made from Gore-Tex have shown a profound waterproof properties that are suitable for stable detection of nitroaromatic explosives such as TNT and 2,4-dinitrotoluene (DNT) in both the liquid- and gas phases [73]. The ability to print electrodes in multiplexed arrays has enabled the fabrication of multiplexed electrochemical sensor [102].

In addition to the application for gas-based detection of explosives, the stretchable biosensors have shown their ability in diagnosis of security threats and environmental contaminants in aqueous solutions [101, 103]. For instance, isolation and transport of sea water samples to the laboratories for further analysis is costly, labour-intensive, and time-consuming, indicating the need for portable devices for on-site analysis of water samples [101]. Moreover, in military activities with a requirement of hands-free operation, a built-in biosensor facilitates their performance and provides the user with a real-time assessment of the environment. For such applications, neoprene-based electrochemical electrodes have been integrated into flexible substrates to provide effective wearable electrochemical-based biosensors for on-site assessment of the presence of contaminants and hazards, such as copper and phenolic pollutants [101]. Toward achieving a fully integrated electrochemical portable biosensor, integration with miniaturized potentiostats [103] has been utilized to detect phenols with the possibility of visual indication and warning system alarming the divers and surfers about toxic level of the environment. The future application of these fully miniaturized sensors is to integrate them into suits worn by divers, surfers, and swimmers, for continuous monitoring and security hazards.

3 Emerging Applications

3.1 Wound Healing

A large number of compromised and chronic non-healing wounds in our ageing population have imposed a significant health and economic burden [104]. This problem is intensified when combined with other complications such as diabetes, vascular disease, and long-term immobility [105]. The current trend in the global wound care market is now expected to raise from $15.6 billion in 2014 to $18.3

billion in 2019 [104]. Therefore, developing novel therapeutic approaches for treating skin wounds has been one of the most active areas of research during the past few decades. The existing approaches for management of skin wound vary from simple coverage of the wound with a sterile patch [106] to smart wound patches capable of measuring the wound environment [81], and releasing the drug on demand [21]. An ideal wound patch should be oxygen permeable, maintain the moist environment of the wound, prevent bacteria and fungi growth, and cover the entire wound area without imposing discomfort and irritation.

Quantitative monitoring of the wound environment during the healing process is of great importance in effective management of the wounds [15, 107, 108]. For example, the ability to measure pH of the wound provides useful information about the stage of the healing process and the possibility of bacterial infection. Another parameter that affects the wound healing process is hydration [107], as the ability to control the moisture balance at the wound surface, central for a successful healing. The hydration of the wound surface can be assessed by measuring the wound temperature and humidity sensors [108]. Measurement of other important markers such as oxygen level, cytokine release, the level of hormones and enzymes at the wound site also provides useful information about the healing process [85].

Multifunctional skin-like electronics capable of covering the wound area have been recently emerged as a promising approach for wound management applications [16, 108]. These electronic systems are composed of metallic microstrips that precisely measure the temperature and thermal conductivity of the wound (Fig. 16.5a, b) [108]. The material characteristics and mechanical properties of the substrate that holds the electronic circuit can be tuned to comfortably cover the wounded area. Electronic devices can also be fabricated in the form of sutures. These "smart" sutures were instrumented with high quality inorganic and biocompatible semiconductors to perform sophisticated sensing and therapeutic actions (Fig. 16.5c, d) [16]. The materials used in these sutures can be biodegradable or permanent [109], while the sutures can be loaded with drugs, growth factors, or stem cells for promoting the wound healing process [110]. Smart sutures can be assembled into complex mesh-like patches using textile technologies for treatment of large-scale wounds.

One of the main design parameters for a wound dressing is the ability to maintain the wound in a moist environment [111]. Hydrogels are 3D polymeric materials containing high water content, and thus are suitable materials for wound dressings [112]. Hydrogels can be impregnated with drugs and growth factors to modulate the vascular formation and also inhibit bacterial growth at the wound site. In addition, other biocompatible materials such as glycerol can improve the flexibility of the hydrogel-based materials and improve their ability to maintain water content over long periods of time [21]. Conductive materials such as gold can be patterned on hydrogel substrates to create electronic circuits for sensing and actuating purposes. For example, a heating element was patterned on an alginate/glycerol sheet using screen printing method to stimulate thermo-responsive drug-loaded microparticles that were embedded within the alginate sheet (Fig. 16.5e, f). This system can be used for on-demand release of antibiotics and growth factors in a programmed manner.
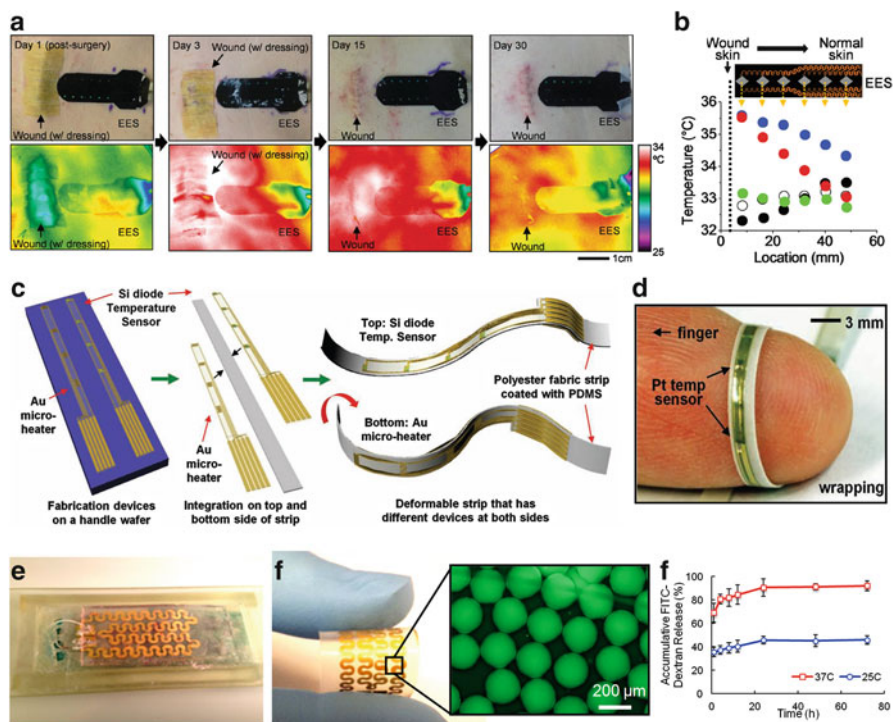
Fig. 16.5 Flexible electronics for wound healing application. (**a**) An epidermal electronics system capable of measuring the wound temperature during the healing process. Infrared images show the temperature distribution over 30 days. (**b**) Temperature distribution measured by the flexible electronic system. Reprinted with permission from [108]. Copyright 2014 by WILEY-VCH Verlag GmbH & Co. KGaA. (**c**) Smart sutures made from strips of biocompatible semiconductors on a polyester fabric. (**d**) Smart suture measuring wrapped around a finger and the temperature. Reprinted with permission from [16]. Copyright 2012 by WILEY-VCH Verlag GmbH & Co. KGaA. (**e**) An integrated hydrogel-based flexible electronic system capable of releasing drugs and small molecules on demand. (**f**) A flexible hydrogel-based substrate with encapsulate thermo-responsive microparticles (*inset*). Release profile of the particles at different temperatures. Reprinted with permission from [21]. Copyright 2015 by WILEY-VCH Verlag GmbH & Co. KGaA

3.2 Implantable Devices and Surgical Tools

Implantable devices are battery-powered electronics that are placed into the body for diagnostics and therapeutic applications. These devices are able to measure several body parameters such as temperature, oxygen content, and pH level, and determine electrophysiological mapping of heart and brain [113]. Current implantable electronics including pacemakers, deep brain stimulators, and epilepsy management devices are composed of electrodes that are in contact with the targeted tissue and a control unit system that monitors each electrode individually [14, 114]. These electrodes are fabricated on flexible substrates such as silicon-based materials [114] or directly patterned on the tissue by a sacrificial substrate that was dissolvable
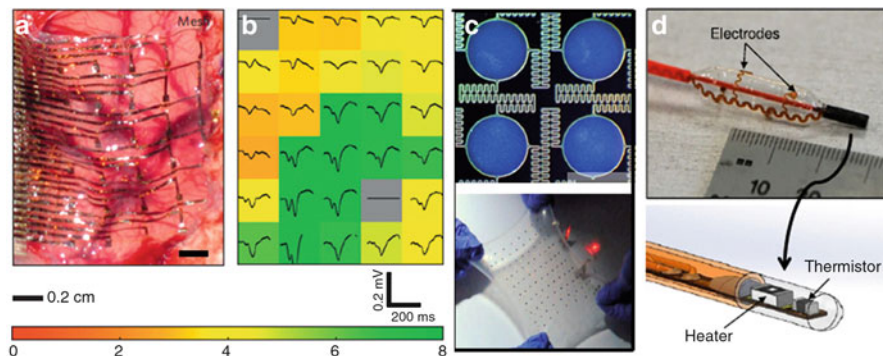
Fig. 16.6 Flexible implantable devices and surgical tools. (**a**) Electrode arrays implanted directly onto a soft brain tissue. (**b**) Measured brain activity signals using the flexible electrode array. Reprinted with permission from [14]. Copyright 2012 by Macmillan Publishers Limited. (**c**) Fractal-like design for fabrication of stretchable batteries. Reprinted with permission from [120]. Copyright 2013 by Nature Publishing Group. (**d**) A balloon catheter system with integrated electrodes for measuring blood flow in real-time and to perform ablation. Reprinted with permission from [19]. Copyright 2015 by Elsevier

in vivo (Fig. 16.6a) [14]. The advantages of these electronic systems are their ability to conform on dynamic and 3D surfaces such as heart or brain and provide electrophysiological mapping of the tissues by means of an array of micropatterned electrodes (Fig. 16.6b) [14]. For example, 2016 silicon nanomembrane transistors were fabricated on a thin plastic sheet and implanted on the surface of a beating porcine heart to record its electrophysiological signals [114]. Unlike conventional electrodes, this technology did not require repositioning of electrodes, and therefore provided a high-resolution and real-time map of the heart.

Miniaturized integrated batteries are essential to supply the power required for long-term function of the implantable devices [115]. The batteries should be able to generate high energy or power densities in a small footprint. Two-dimensional (2D) thin-film batteries supplied high power densities; however, they need large footprints to deliver high energies required for activating some of the electronic components [115]. It has been shown that colloidally templated 3D mesostructured electrodes could improve the efficacy of power density production by delivering significant energy due to high surface to volume ratios [116]. This technology can take advantage of 3D holographic patterning techniques to allow precise control over the microstructure of the electrodes. Stretchable batteries can be fabricated from stretchable electrodes using wavy and wrinkled designs from conductive materials such as polypyrrole [117]. The fabrication methods are similar to the methods used for patterning of sensors on elastomeric substrates. Other design configurations such as origami [118], serpentine [119], and fractal-like [120] patterns have also been recently developed to achieve high power densities and larger stretchability (Fig. 16.6c).

Catheter-based diagnostic and treatment approaches are non-invasive methods that rely on the local delivery of sensing elements and therapeutic payloads. Catheters, instrumented with recording and stimulating electrodes, are flexible and therefore can interface with the soft tissues within the human body. Balloon-based catheters with multiple electrodes have been developed to map the electrical activity of the heart and perform ablation therapy [13]. The fabrication strategy includes patterning of ultrathin conductive layers and inorganic nanomaterials on inflated elastomeric balloon substrates using plasma treatment, transfer printing, and interfacial adhesives [13, 121, 122]. Another application of balloon-based catheters is to manage vascular diseases such as coronary artery diseases using an approach referred as percutaneous transluminal coronary angioplasty [121]. These catheter-based methods are usually combined with drug-eluting stents for local delivery of therapeutic agents [123]. To improve the functionality of these balloons, blood sensors and actuators have also been integrated into the electronic system to monitor the blood vessel health in real time (Fig. 16.6d) [19]. These sophisticated platforms enable performing surgeries faster with less instrumentation and higher vascular intervention efficiency.

3.3   Wearable Devices

Wearable systems are defined as a new class of the health monitoring and therapeutic platform that can be embedded into the user clothing or accessory [124]. These systems enable collection of physiological information over a long term without introducing inconvenience to the patient [124]. This ambulatory technology allows real-time monitoring of the key markers in patients with chronic conditions, not necessarily in the hospital. Miniaturization, lightweight, and conformability of the components are the key requirements of wearable systems. These systems contain different electronic and sensing components integrated into a single flexible platform such as bandage and wristband [125, 126].

In the past few years, several wearable devices have been introduced for monitoring of the heart rate [127] [128], the blood pressure [129, 130], electrocardiography [131], and physical activity [39, 132]. In a notable study, a wearable strain sensor was fabricated from aligned thin films of CNTs on PDMS as an elastic substrate [39, 133]. CNTs were aligned perpendicular to the strain axis. This system was able to measure strains up to 280 % and was used to assimilate various dynamic motion of the human (Fig. 16.7a–c).

To enhance the adaptability on a rough skin surface, the same group developed a dry adhesive patch for in vitro diagnostics [134]. They presented a simple method for fabricating an enhanced dry adhesive skin patch by utilizing modulus-tenable composite micropillars made of stiff and soft PDMS materials (Fig. 16.7d–f). The method consisted of direct replica moulding of rigid bottom micropillars (curing agent 15 wt%) and selective inking of soft tip layer (curing agent 5 wt%). Such monolithically integrated composite micropillars provided a large normal adhesion
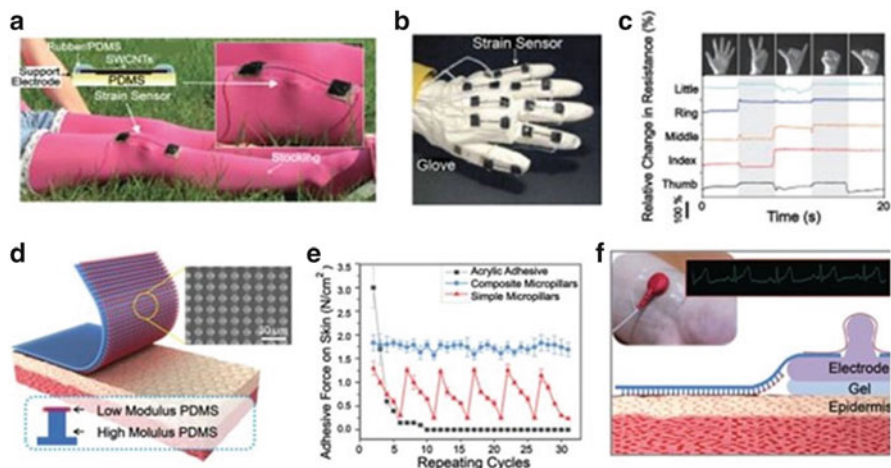
Fig. 16.7 Flexible wearable devices. (**a–c**) Wearable stretchable human-motion detector, strain-gauge sensors and signal patterns of attachment on knee and a glove. Reprinted with permission from [133]. Copyright 2011 by Nature Publishing Group. (**b**) Enhanced skin adhesive patch with composite micropillars with (**d**) a schematic illustration and scanning electron microscopy (SEM) image of heterogeneous micropillars. (**e**) Plot of the enhancement of adhesion, and (**f**) ECG signal measurement with an electrode attached on the skin. Reprinted with permission from [134]. Copyright 2013 by WILEY-VCH Verlag GmbH & Co. KGaA

force up to 1.8 N cm^2 (maximum: 2 N cm^2) on the human skin as well as high durability (30 cycles) without notable degradations. Using the composite micropillars electrocardiograms (ECGs) were successfully recorded in real time for the time period of 48 h with minimal side effects [134].

Wearable chemical sensors are also recently introduced. Smart bandage with embedded oxygen sensors and readout electronics allows continuous monitoring of chronic wound oxygenation (Fig. 16.7g, h) [135, 136]. In addition, electrochemical sensors with screen-printed electrodes on underwater garments are proposed for monitoring of environmental pollutants [137].

Another important application of the wearable devices is the electronic artificial skin (e-skin) [8, 10, 35, 40, 41, 138]. Development of an artificial skin capable of detecting external environmental stimulus is a great interest for the humanoid robotics and prosthetics [10, 35, 41]. E-skin should be implemented on a highly flexible and soft substrate to mimic the human skin. Currently, there has been a great deal of research focusing on the development of e-skin including physical sensors, such as pressure and temperature sensors. The e-skin design generally contains two components, an active layer containing organic transistors, switches, etc. covers the flexible substrate. The second layer includes physical sensors to measure pressure and temperature, and to collect the environmental information. Information of the second layer is read by the first layer as an active layer [41]. Someya et al. presented a net shape e-skin including flexible pressure and temperature sensors implemented

on a polyamide substrate in contact with organic transistors [10, 11]. The platform was conformable to 3D surfaces and enabled the e-skin to extend by 25 %.

3.4 Point-of-Care Devices

The cost of public healthcare is significantly increasing in the modern societies. Point-of-care devices and portable instruments can play essential role to shift from expensive healthcare to low cost care [139]. These devices can be used in diagnosis, prevention, and treatment of diseases to improve the clinical outcome in health care. Such devices are expected to be rapid, cost-effective, portable, multifunctional, with integrated sample preparation, detection and processing systems. Stretchable sensors with high flexibility and sensitivity, and possibility of real-time detection can be miniaturized and function under considerable physical deformation with the potential application in development of point-of-care devices.

A flexible paper-based electrochemical portable biosensor was fabricated based on free-standing graphene coated with nanocomposites of Pt/Au alloy and MnO_2 for the purpose of non-enzymatic amperometric detection of biomolecules [140] (Fig. 16.8a, b). The fabricated sensor offered enhanced sensing with high sensitivity, low detection limit, high reproducibility and stability under mechanical stress, with an acceptable selectivity. An impedance-based electrochemical biosensor was further fabricated by growing CNTs on flexible polyimide substrate, where the CNTs surface was coated with anti-human serum albumin (AHSA) for specific detection of HSA and with the potential application in the implant-based detection [141] (Fig. 16.8c).

Another flexible electrochemical biosensor was made of a graphene paper covered with Au/Pt core–shell nanoparticles for real-time monitoring the reactive oxygen species (ROS) and reactive nitrogen species (RNS), particularly nitric oxide [142]. The integration of metal nanostructures and graphene paper into functional electrodes was optimized to achieve a high electrocatalytic activity and sensitivity with a wide linear range and a low detection limit. Other flexible-based substrate electrochemical biosensors were developed by printing silver electrodes over PET substrate to detect D-proline, sarcosine, cadmium sulphide (CdS), and potassium chloride (KCl) [143]. Hg^{2+} ions exist in low concentrations in biological and environmental samples have extreme toxicity effects for both human health and environment, and may lead to fatal diseases like minamata disease, cyanosis, nephrotic syndrome pulmonary, or edema [144]. A liquid-ion gated FET was integrated into PEN-based flexible graphene apta-sensor with excellent mechanical properties for highly sensitive and selective detection of Hg (Fig. 16.8d) [73]. The sensor exhibited a rapid response of less than 1 s when the Hg^{2+} ion concentration was altered. Additionally, its performance was selective in a mixed solution containing several other metal ions as well as the real blood samples. The sensor was rapidly considered as an effective alternative for silicon-based electronics. Lactic acid is another agent important for environmental monitoring and
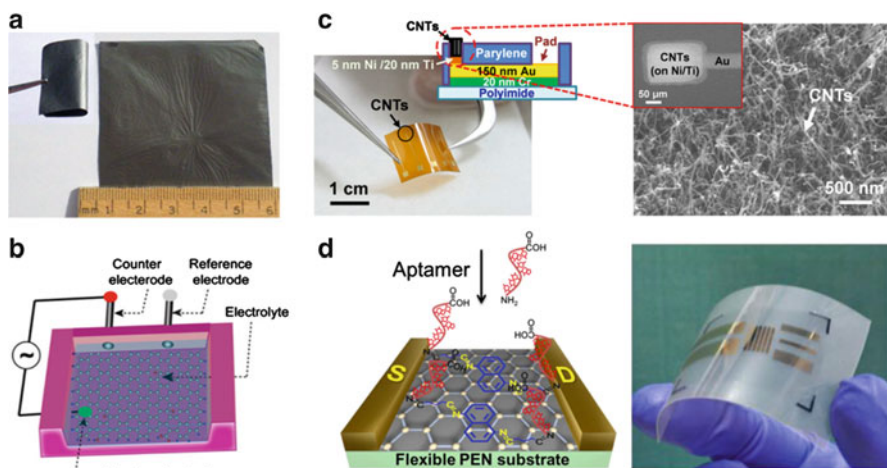
Fig. 16.8 Flexible sensors for point-of-care diagnosis applications. (**a**) Graphene nanosheets. (**b**) The schematic representation of the electrochemical set-up. Reprinted with permission from [140]. Copyright 2013 by Elsevier. (**c**) Flexible CNT electrode with a SEM image of the electrode coated with CNTs. Reprinted with permission from [141]. Copyright 2013 by Elsevier. (**d**) Flexible and transparent graphene-based biosensor deposited the gold electrodes [73]. Reprinted with permission from [73]. Copyright 2013 by ACS Publications

for clinical assessment. Therefore, the development of a flexible biosensor for POC-based detection of lactate is essential. Recently, a flexible graphene-based biosensor was developed to detect lactate under different mechanical bending conditions with the sensitivity of 0.1–20 μM. The performance of the sensor was shown dependent on the bending angle and the number of bending cycles [145].

Stretchable biosensors have also been used for early detection of cancer biomarkers from blood samples [146]. A FET-type apta-sensor was fabricated based on N-doped graphene using conducting polymers with heteroatoms as the carbonization precursor over a flexible substrate. The PPy-NDFLG conjugated with antivascular endothelial growth factor (VEGF) RNA aptamer was embedded into a liquid-ion gated FET sensing area to detect VEGF as angiogenesis biomarker useful for the detection of metastasis and growth of human tumour. The sensor demonstrated enhanced reusability and mechanical bendability required for a sensor for point-of-care application.

Wireless communication is an inevitable component of point-of-care devices for long-term monitoring of physiological signals. A wireless biosensor printed on a stretchable substrate with integrated antennas and interconnections between components was utilized to monitor biomarkers. A wireless biosensor with inkjet conductors printed over polyurethane substrate was also successfully utilized to detect organophosphate (OP) and p-nitrophenol hydrolysis with the measured reading range of 14 m distance from the body [147].

4 Summary and Conclusions

The advances highlighted in this chapter shed light on the exciting opportunities offered by the flexible electronics technology. These directions emerge from the unique characteristics of flexible electronic systems that allow integrating electronic components into soft tissues with irregular surfaces. Novel fabrication methods combined with advanced stretchable biomaterials enabled developing complex electronic devices that can monitor physical and biological signals within the body or from outer layers of skin and transport these valuable information to central unit for further analysis. Integration of the sensors with other electronic components such as power supply and communication elements is essential for prolong measurement of the physical markers. An important factor in designing such systems is to ensure long-live and robust electrical contacts that do not cause discomfort and irritation for the tissue. More advanced flexible biosensors have been developed that can be mounted on medical catheters for minimally invasive detection of biomarkers. These sensors can be delivered to the site of interest intravascularly, measure the markers and if needed release proper treatment locally. Another exciting application of flexible electronics is the development of smart dressings capable of on-demand release of antibacterial agents and growth factors at the site of injury. Flexible electronics can also be wearable to collect the physiological information over a long term without the patient inconvenience. This ambulatory technology allows real-time monitoring of the key markers in patients with chronic conditions, not necessarily in the hospitals. A major application of flexible electronics is in development of point-of-care devices and portable instruments that are essential for improving the quality of public health in low resource settings. These devices can be used in diagnosis, prevention, and treatment of diseases to improve the clinical outcome in health care. All together, flexible electronics technology holds a great promise for a wide range of biomedical applications. However, advances in the integration of wireless communication systems, substrate materials with tuneable degradation and mechanical properties, and manufacturing technologies that allow large-scale fabrication of electronic devices are necessary.

References

1. Crabb R, Treble F (1967) Thin silicon solar cells for large flexible arrays. Nature 213: 1223–1224
2. Shahrjerdi D et al (2013) Flexible solar cells: ultralight high-efficiency flexible InGaP/(In) GaAs tandem solar cells on plastic (Adv. Energy Mater. 5/2013). Adv Energy Mater 3(5):542
3. Romeo A et al (2006) High-efficiency flexible CdTe solar cells on polymer substrates. Sol Energy Mater Sol Cells 90(18):3407–3415
4. Nathan A et al (2012) Flexible electronics: the next ubiquitous platform. Proc IEEE 100(Special Centennial Issue):1486–1517
5. Qi Y et al (2010) Piezoelectric ribbons printed onto rubber for flexible energy conversion. Nano Lett 10(2):524–528

6. Wang ZL (2010) Toward self-powered sensor networks. Nano Today 5(6):512–514
7. Manekkathodi A et al (2010) Direct growth of aligned zinc oxide nanorods on paper substrates for low-cost flexible electronics. Adv Mater 22(36):4059–4063
8. Lipomi DJ et al (2011) Skin-like pressure and strain sensors based on transparent elastic films of carbon nanotubes. Nat Nanotechnol 6(12):788–792
9. Rogers JA et al (2001) Paper-like electronic displays: large-area rubber-stamped plastic sheets of electronics and microencapsulated electrophoretic inks. Proc Natl Acad Sci 98(9):4835–4840
10. Someya T et al (2005) Conformable, flexible, large-area networks of pressure and thermal sensors with organic transistor active matrixes. Proc Natl Acad Sci USA 102(35):12321–12325
11. Someya T et al (2004) A large-area, flexible pressure sensor matrix with organic field-effect transistors for artificial skin applications. Proc Natl Acad Sci USA 101(27):9966–9970
12. Pang C et al (2012) A flexible and highly sensitive strain-gauge sensor using reversible interlocking of nanofibres. Nat Mater 11(9):795–801
13. Kim D-H et al (2012) Flexible and stretchable electronics for biointegrated devices. Annu Rev Biomed Eng 14:113–128
14. Kim D-H et al (2010) Dissolvable films of silk fibroin for ultrathin conformal bio-integrated electronics. Nat Mater 9(6):511–517
15. Kim D-H et al (2011) Epidermal electronics. Science 333(6044):838–843
16. Kim DH et al (2012) Thin, flexible sensors and actuators as 'instrumented' surgical sutures for targeted wound monitoring and therapy. Small 8(21):3263–3268
17. Kim T et al (2012) Large-scale graphene micropatterns via self-assembly-mediated process for flexible device application. Nano Lett 12(2):743–748
18. Lee J-Y et al (2009) Flexible electrochemical biosensors based on O2 plasma functionalized MWCNT. Thin Solid Films 517(14):3883–3887
19. Klinker L et al (2015) Balloon catheters with integrated stretchable electronics for electrical stimulation, ablation and blood flow monitoring. Extreme Mech Lett 3:45–54
20. Santini JT Jr, Hutchinson CE (2003) Implantable drug delivery stents. Google Patents
21. Bagherifard S et al (2015) Dermal patch with integrated flexible heater for on demand drug delivery. Adv Healthcare Mater 5(1):175–184
22. Choi WM et al (2007) Biaxially stretchable "wavy" silicon nanomembranes. Nano Lett 7(6):1655–1663
23. Kim D-H et al (2008) Materials and noncoplanar mesh designs for integrated circuits with linear elastic responses to extreme mechanical deformations. Proc Natl Acad Sci 105(48):18675–18680
24. Huang K, Peumans P (2006) Stretchable silicon sensor networks for structural health monitoring. In: Smart structures and materials. International Society for Optics and Photonics
25. Hasan M et al (2010) Low temperature aluminum oxide gate dielectric on plastic film for flexible device application. Jpn J Appl Phys 49(5S1):05EA01
26. Duan X et al (2003) High-performance thin-film transistors using semiconductor nanowires and nanoribbons. Nature 425(6955):274–278
27. McAlpine MC et al (2003) High-performance nanowire electronics and photonics on glass and plastic substrates. Nano Lett 3(11):1531–1535
28. Gelinck GH et al (2004) Flexible active-matrix displays and shift registers based on solution-processed organic transistors. Nat Mater 3(2):106–110
29. Bae S et al (2010) Roll-to-roll production of 30-inch graphene films for transparent electrodes. Nat Nanotechnol 5(8):574–578
30. Ahn J-H et al (2006) Heterogeneous three-dimensional electronics by use of printed semiconductor nanomaterials. Science 314(5806):1754–1757
31. Hines D et al (2007) Transfer printing methods for the fabrication of flexible organic electronics. J Appl Phys 101(2):024503
32. Najafabadi AH et al (2014) Biodegradable nanofibrous polymeric substrates for generating elastic and flexible electronics. Adv Mater 26(33):5823–5830

33. Huang D et al (2003) Plastic-compatible low resistance printable gold nanoparticle conductors for flexible electronics. J Electrochem Soc 150(7):G412–G417
34. Fan F-R et al (2012) Transparent triboelectric nanogenerators and self-powered pressure sensors based on micropatterned plastic films. Nano Lett 12(6):3109–3114
35. Mannsfeld SC et al (2010) Highly sensitive flexible pressure sensors with microstructured rubber dielectric layers. Nat Mater 9(10):859–864
36. Muth JT et al (2014) Embedded 3D printing of strain sensors within highly stretchable elastomers. Adv Mater 26(36):6307–6312
37. Zhu S et al (2013) Ultrastretchable fibers with metallic conductivity using a liquid metal alloy core. Adv Funct Mater 23(18):2308–2314
38. Boley JW et al (2014) Direct writing of gallium-indium alloy for stretchable electronics. Adv Funct Mater 24(23):3501–3507
39. Yamada T et al (2011) A stretchable carbon nanotube strain sensor for human-motion detection. Nat Nanotechnol 6(5):296–301
40. Hammock ML et al (2013) 25th anniversary article: the evolution of electronic skin (E-Skin): a brief history, design considerations, and recent progress. Adv Mater 25(42):5997–6038
41. Takei K et al (2010) Nanowire active-matrix circuitry for low-voltage macroscale artificial skin. Nat Mater 9(10):821–826
42. Niu X et al (2007) Characterizing and patterning of PDMS-based conducting composites. Adv Mater-Deerfield Beach Then Weinheim- 19(18):2682
43. Sun JY et al (2014) Ionic skin. Adv Mater 26(45):7608–7614
44. Chen B et al (2014) Highly stretchable and transparent ionogels as nonvolatile conductors for dielectric elastomer transducers. ACS Appl Mater Interfaces 6(10):7840–7845
45. Krebs FC, Fyenbo J, Jørgensen M (2010) Product integration of compact roll-to-roll processed polymer solar cell modules: methods and manufacture using flexographic printing, slot-die coating and rotary screen printing. J Mater Chem 20(41):8994–9001
46. Ladd C et al (2013) 3D printing of free standing liquid metal microstructures. Adv Mater 25(36):5081–5085
47. Park Y-L et al (2010) Hyperelastic pressure sensing with a liquid-embedded elastomer. J Micromech Microeng 20(12):125029
48. Rogers JA, Someya T, Huang Y (2010) Materials and mechanics for stretchable electronics. Science 327(5973):1603–1607
49. Tao H et al (2011) Metamaterials on paper as a sensing platform. Adv Mater 23(28):3197–3201
50. Hyun WJ, Park OO, Chin BD (2013) Foldable graphene electronic circuits based on paper substrates. Adv Mater 25(34):4729–4734
51. Felmet K, Loo Y-L, Sun Y (2004) Patterning conductive copper by nanotransfer printing. Appl Phys Lett 85(15):3316–3318
52. Hwang S-W et al (2012) A physically transient form of silicon electronics. Science 337(6102):1640–1644
53. Sekitani T et al (2010) Flexible organic transistors and circuits with extreme bending stability. Nat Mater 9(12):1015–1022
54. Park M et al (2013) Micropatterned stretchable circuit and strain sensor fabricated by lithography on an electrospun nanofiber mat. ACS Appl Mater Interfaces 5(17):8766–8771
55. Bettinger CJ, Bao Z (2010) Organic thin-film transistors fabricated on resorbable biomaterial substrates. Adv Mater 22(5):651–655
56. Tamayol A et al (2013) Fiber-based tissue engineering: progress, challenges, and opportunities. Biotechnol Adv 31(5):669–687
57. Akbari A, Akbari M, Hill RJ (2013) Effective thermal conductivity of two-dimensional anisotropic two-phase media. Int J Heat Mass Transf 63:41–50
58. Collins A, Cosh J (1970) Temperature and biochemical studies of joint inflammation. A preliminary investigation. Ann Rheum Dis 29(4):386
59. Brown DL (1999) Device and method for locating inflamed plaque in an artery. Google Patents

60. Savla U, Olson LE, Waters CM (2004) Mathematical modeling of airway epithelial wound closure during cyclic mechanical strain. J Appl Physiol 96(2):566–574
61. Lichtenwalner DJ, Hydrick AE, Kingon AI (2007) Flexible thin film temperature and strain sensor array utilizing a novel sensing concept. Sens Actuators A Phys 135(2):593–597
62. Briand D et al (2011) Making environmental sensors on plastic foil. Mater Today 14(9):416–423
63. Xiao SY et al (2008) A novel fabrication process of MEMS devices on polyimide flexible substrates. Microelectron Eng 85(2):452–457
64. Xiao S-Y et al (2005) A temperature sensor array based on flexible MEMS skin technology [J]. Opt Precis Eng 6:008
65. Sibinski M, Jakubowska M, Sloma M (2010) Flexible temperature sensors on fibers. Sensors 10(9):7934–7946
66. Shih W-P et al (2010) Flexible temperature sensor array based on a graphite-polydimethylsiloxane composite. Sensors 10(4):3597–3610
67. Karimov KS, Chani MTS, Khalid FA (2011) Carbon nanotubes film based temperature sensors. Physica E Low-dimensional Syst Nanostruct 43(9):1701–1703
68. Karimov KS et al (2015) Temperature sensor based on composite film of vanadium complex (VO2 (3-fl)) and CNT. J Semicond 36(7):073004
69. Frutiger A et al (2015) Capacitive soft strain sensors via Multicore–Shell fiber printing. Adv Mater 27(15):2440–2446
70. Ridzuan NA, Masuda S, Miki N (2012) Flexible capacitive sensor encapsulating liquids as dielectric with a largely deformable polymer membrane. Micro Nano Lett 7(12):1193–1196
71. Sekitani T et al (2005) Bending experiment on pentacene field-effect transistors on plastic films. Appl Phys Lett 86(7):73511–74100
72. Sekitani T et al (2009) Stretchable active-matrix organic light-emitting diode display using printable elastic conductors. Nat Mater 8(6):494–499
73. Windmiller JR, Wang J (2013) Wearable electrochemical sensors and biosensors: a review. Electroanalysis 25(1):29–46
74. Cai J et al (2009) Flexible thick-film electrochemical sensors: impact of mechanical bending and stress on the electrochemical behavior. Sens Actuators B Chem 137(1):379–385
75. Shao Y et al (2010) Graphene based electrochemical sensors and biosensors: a review. Electroanalysis 22(10):1027–1036
76. Boukhvalov D, Katsnelson M (2009) Chemical functionalization of graphene. J Phys Condens Matter 21(34):344205
77. Wang Y et al (2010) Nitrogen-doped graphene and its application in electrochemical biosensing. ACS Nano 4(4):1790–1798
78. Bagri A et al (2010) Stability and formation mechanisms of carbonyl-and hydroxyl-decorated holes in graphene oxide. J Phys Chem C 114(28):12053–12061
79. Gethin G (2007) The significance of surface pH in chronic wounds. Wounds UK 3(3):52
80. Huang X et al (2014) Stretchable, wireless sensors and functional substrates for epidermal characterization of sweat. Small 10(15):3083–3090
81. Tamayol A et al Flexible pH-sensing hydrogel fibers for epidermal applications (5):711–719
82. Cramer W, Butler W (1969) Potentiometric titration of the fluorescence yield of spinach chloroplasts. Biochim Biophys Acta Bioenergetics 172(3):503–510
83. Huang W-D et al (2011) A flexible pH sensor based on the iridium oxide sensing film. Sensors Actuators A Phys 169(1):1–11
84. Kurzweil P (2009) Metal oxides and ion-exchanging surfaces as pH sensors in liquids: state-of-the-art and outlook. Sensors 9(6):4955–4985
85. Dargaville TR et al (2013) Sensors and imaging for wound healing: a review. Biosens Bioelectron 41:30–42
86. Sharp D (2013) Printed composite electrodes for in-situ wound pH monitoring. Biosens Bioelectron 50:399–405
87. Kaempgen M, Roth S (2006) Transparent and flexible carbon nanotube/polyaniline pH sensors. J Electroanal Chem 586(1):72–76

88. Ferrer-Anglada N, Kaempgen M, Roth S (2006) Transparent and flexible carbon nanotube/polypyrrole and carbon nanotube/polyaniline pH sensors. Phys Status Solidi B 243(13):3519–3523

89. Korostynska O et al (2007) Review on state-of-the-art in polymer based pH sensors. Sensors 7(12):3027–3042

90. Madou MJ, Morrison SR (2012) Chemical sensing with solid state devices. Elsevier, Amsterdam

91. Chuang M-C et al (2010) Flexible thick-film glucose biosensor: influence of mechanical bending on the performance. Talanta 81(1):15–19

92. Mitsubayashi K et al (2003) Optical-transparent and flexible glucose sensor with ITO electrode. Biosens Bioelectron 19(1):67–71

93. Mani V, Devadas B, Chen S-M (2013) Direct electrochemistry of glucose oxidase at electrochemically reduced graphene oxide-multiwalled carbon nanotubes hybrid material modified electrode for glucose biosensor. Biosens Bioelectron 41:309–315

94. Urban G et al (1992) Miniaturized multi-enzyme biosensors integrated with pH sensors on flexible polymer carriers for in vivo applications. Biosens Bioelectron 7(10):733–739

95. Kagie A et al (2008) Flexible rolled thick-film miniaturized flow-cell for minimally invasive amperometric sensing. Electroanalysis 20(14):1610–1614

96. Morris D et al (2009) Bio-sensing textile based patch with integrated optical detection system for sweat monitoring. Sens Actuators B Chem 139(1):231–236

97. Yang Y-L et al (2010) Thick-film textile-based amperometric sensors and biosensors. Analyst 135(6):1230–1234

98. Li C, Han J, Ahn CH (2007) Flexible biosensors on spirally rolled micro tube for cardiovascular in vivo monitoring. Biosens Bioelectron 22(9):1988–1993

99. Li C et al (2008) A flexible polymer tube lab-chip integrated with microsensors for smart microcatheter. Biomed Microdevices 10(5):671–679

100. Wang J (2004) Microchip devices for detecting terrorist weapons. Anal Chim Acta 507(1):3–10

101. Wang J (2007) Electrochemical sensing of explosives. In: Counterterrorist detection techniques of explosives. pp 91–108

102. Chuang MC et al (2010) Textile-based electrochemical sensing: effect of fabric substrate and detection of nitroaromatic explosives. Electroanalysis 22(21):2511–2518

103. Windmiller JR et al (2011) Bioelectronic system for the control and readout of enzyme logic gates. Sens Actuators B Chem 155(1):206–213

104. Sen CK et al (2009) Human skin wounds: a major and snowballing threat to public health and the economy. Wound Repair Regen 17(6):763–771

105. Greer N et al (2013) Advanced wound care therapies for nonhealing diabetic, venous, and arterial ulcers: a systematic review. Ann Intern Med 159(8):532–542

106. Bryant RA, Nix DP (2012) Acute & chronic wounds: current management concepts. Elsevier Health Sciences, St. Louis

107. McColl D, Cartlidge B, Connolly P (2007) Real-time monitoring of moisture levels in wound dressings in vitro: an experimental study. Int J Surg 5(5):316–322

108. Hattori Y et al (2014) Multifunctional skin-like electronics for quantitative, clinical monitoring of cutaneous wound healing. Adv Healthcare Mater 3(10):1597–1607

109. Akbari M et al Textile technologies and tissue engineering: a path toward organ weaving, Adv. Healthcare Mater. 2016, (5): 751–766

110. Akbari M et al (2014) Composite living fibers for creating tissue constructs using textile techniques. Adv Funct Mater 24(26):4060–4067

111. Jones A, Vaughan D (2005) Hydrogel dressings in the management of a variety of wound types: a review. J Orthop Nurs 9(Supplement 1):S1–S11

112. Annabi N et al (2014) 25th anniversary article: rational design and applications of hydrogels in regenerative medicine. Adv Mater 26(1):85–124

113. Ortiz-Catalan M et al (2012) On the viability of implantable electrodes for the natural control of artificial limbs: review and discussion. Biomed Eng Online 11:33

114. Viventi J et al (2010) A conformal, bio-interfaced class of silicon electronics for mapping cardiac electrophysiology. Sci Transl Med 2(24):24ra22
115. Patil A et al (2008) Issue and challenges facing rechargeable thin film lithium batteries. Mater Res Bull 43(8):1913–1942
116. Ning H et al (2015) Holographic patterning of high-performance on-chip 3D lithium-ion microbatteries. Proc Natl Acad Sci 112(21):6573–6578
117. Wang C et al (2011) Buckled, stretchable polypyrrole electrodes for battery applications. Adv Mater 23(31):3580–3584
118. Song Z et al (2014) Origami lithium-ion batteries. Nat Commun 5, 3140
119. Kim D et al (2013) Fabrication of a stretchable solid-state micro-supercapacitor array. ACS Nano 7(9):7975–7982
120. Xu S et al (2013) Stretchable batteries with self-similar serpentine interconnects and integrated wireless recharging systems. Nat Commun 4:1543
121. Slepian MJ, Ghaffari R, Rogers JA (2011) Multifunctional balloon catheters of the future. Interv Cardiol 3(4):417–419
122. Lee SP et al (2015) Catheter-based systems with integrated stretchable sensors and conductors in cardiac electrophysiology. Proc IEEE 103(4):682–689
123. Hawkins BM, Hennebry TA (2011) Local paclitaxel delivery for treatment of peripheral arterial disease. Circ Cardiovasc Interv 4(3):297–302
124. Patel S et al (2012) A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 9(12):1–17
125. Heldberg BE et al (2015) Using wearable sensors for semiology-independent seizure detection-towards ambulatory monitoring of epilepsy. In: Engineering in medicine and biology society (EMBC), 2015 37th annual international conference of the IEEE
126. Bawazer LA (2014) Bio focus: healthcare-on-a-patch: responsive wearable materials. MRS Bull 39(06):477
127. Paradiso R, Loriga G, Taccini N (2005) A wearable health care system based on knitted integrated sensors. IEEE Trans Inf Technol Biomed 9(3):337–344
128. Martin T, Jovanov E, Raskovic D (2000) Issues in wearable computing for medical monitoring applications: a case study of a wearable ECG monitoring device. In: The fourth international symposium on wearable computers, IEEE
129. Anliker U et al (2004) AMON: a wearable multiparameter medical monitoring and alert system. IEEE Trans Inf Technol Biomed 8(4):415–427
130. Pandian P et al (2008) Smart Vest: wearable multi-parameter remote physiological monitoring system. Med Eng Phys 30(4):466–477
131. Leonov V, Vullers RJ (2009) Wearable electronics self-powered by using human body heat: the state of the art and the perspective. J Renewable Sustainable Energy 1(6):062701
132. Jung S, Ji T, Varadan VK (2006) Point-of-care temperature and respiration monitoring sensors for smart fabric applications. Smart Mater Struct 15(6):1872
133. Yamada T et al (2011) A stretchable carbon nanotube strain sensor for human-motion detection. Nat Nanotechnol 6(5):296–301
134. Bae WG et al (2013) Enhanced skin adhesive patch with Modulus-Tunable composite micropillars. Adv Healthcare Mater 2(1):109–113
135. Mostafalu P et al (2015) Wireless flexible smart bandage for continuous monitoring of wound oxygenation. IEEE Trans Biomed Circuits Syst 9(5):670–677
136. Mostafalu P et al (2014) Wireless flexible smart bandage for continuous monitoring of wound oxygenation. In: 2014 IEEE biomedical circuits and systems conference (BioCAS), IEEE
137. Malzahn K et al (2011) Wearable electrochemical sensors for in situ analysis in marine environments. Analyst 136(14):2912–2917
138. Schwartz G et al (2013) Flexible polymer transistors with high pressure sensitivity for application in electronic skin and health monitoring. Nat Commun 4:1859
139. Chin CD, Linder V, Sia SK (2012) Commercialization of microfluidic point-of-care diagnostic devices. Lab Chip 12(12):2118–2134

140. Xiao F et al (2013) Growth of coral-like PtAu–MnO$_2$ binary nanocomposites on free-standing graphene paper for flexible nonenzymatic glucose sensors. Biosens Bioelectron 41:417–423
141. Chang Y-T et al (2013) Flexible direct-growth CNT biosensors. Biosens Bioelectron 41:898–902
142. Zan X et al (2013) Freestanding graphene paper decorated with 2D-assembly of Au@ Pt nanoparticles as flexible biosensors to monitor live cell secretion of nitric oxide. Biosens Bioelectron 49:71–78
143. Reddy A et al (2010) Printed electrochemical based biosensors on flexible substrates. In: 2010 IEEE sensors, IEEE
144. Clarkson TW, Magos L, Myers GJ (2003) The toxicology of mercury—current exposures and clinical manifestations. N Engl J Med 349(18):1731–1737
145. Labroo P, Cui Y (2013) Flexible graphene bio-nanosensor for lactate. Biosens Bioelectron 41:852–856
146. Kwon OS et al (2012) Flexible FET-type VEGF aptasensor based on nitrogen-doped graphene converted from conducting polymer. ACS Nano 6(2):1486–1493
147. Sillanpaa H et al (2014) Inkjet printed wireless biosensors on stretchable substrate. In: 2014 international conference on electronics packaging (ICEP), IEEE

Chapter 17
MEMS Devices in Agriculture

Antonio Valente

Abstract This chapter is dedicated to Micro-electro-mechanical Systems (MEMS) devices developed for primary use in agriculture. We can see MEMS devices in ink jet printers (printer heads), automobiles (e.g., airbag accelerometer), projectors (digital micromirror device for DLP projectors), mobile devices (e.g., gyroscopes for smartphones, tablets, etc.), healthcare applications (e.g., lab-on-a-chip for the detection of multiple tropical infectious diseases), among others. Although it is expected to grow, the use of MEMS devices used in agriculture still comes down to a few cases in research centers. However, due to demand for improved agricultural processes and the future widespread use of the Internet of Things (IoT), a high demand for small size, low cost, low power, and easily mass produced devices is expected. This context suggests the use of MEMS devices for both sensing elements, and for energy harvesters. In this chapter, only the sensor elements whose major use is agriculture will be addressed. From this perspective, the main parameters used in agriculture will be addressed taking into account research and development ever held in MEMS devices for measuring these parameters. These key parameters are grouped into classes: environment, soil, agricultural crops, and livestock. For almost all of the parameters shown, MEMS devices showed encouraging results. Most work with MEMS for agriculture has been done in laboratories so far. However, transitioning to field applications seems feasible. Potential advantages of MEMS are: small size, economical production (specially in large scale), built-in electronics (for auto-calibration, self-testing, digital compensation, and digital communications), and low power consumption—ideal for the use in Precision Agriculture complemented by Internet of Things. In short, this chapter will allow researchers developing MEMS devices to have a knowledge of what has already been developed for agriculture and to have an idea of future needs in this field.

Keywords MEMS • Sensors • Agriculture sensors • Soil sensors • Crop sensors • Livestock sensors • Environment sensor

A. Valente (✉)
INESC TEC (formerly INESC Porto) and School of Science and Technology,
UTAD University, Vila Real, Portugal
e-mail: avalente@utad.pt

1   Introduction

The term "Micro-electro-mechanical Systems" (MEMS) was first introduced in 1986. Since then millions of devices were made and commercialized in the world. Nowadays, we can see MEMS devices in inkjet printers (printer heads), accelerometers and gyroscopes (modern cars airbags, cell phones, drones, game controllers, digital cameras, gadgets, among others), microphones, pressure sensors, media projectors (DLP), data communication (optical switch), bio-MEMS applications in health care (Lab-on-Chip), and many other examples. It is possible to see in this short enumeration that the main areas are automotive, consumer goods, and health. The use of MEMS in agriculture is almost by secondary appearance (e.g., drones used in agriculture have MEMS accelerometers).

With the introduction of precision agriculture (PA) concept—management within fields by zones rather than whole fields for increased profitability—and the new communication era with Internet-of-Things (IoT) the use of MEMS devices in agriculture is increasing. Someone said [54],

> Through the ages agriculture production systems have benefited from the incorporation of technological advances primarily developed for other industries. The industrial age brought mechanization and synthesized fertilizers, the technological age offered genetic engineering and now the information age brings the potential for Precision Agriculture.

and MEMS devices are definitively part of the new information age.

MEMS market is expected to almost double from a \$11 billion business in 2014 to \$21 billion in 2020 (a two-digit annual growth, Fig. 17.1).

> If the Internet of Things (IoT) is going to drive the next round of electronics industry growth, it will depend in large part on the MEMS and sensor technology that will enable all those smart objects to interact with the real world. But the ramp of new MEMS designs to volume production may take too long and cost too much to meet IoT market expectations, unless the industry figures out ways to accelerate MEMS development [17].

In this report from Doe [17] in EETimes (http://www.eetimes.com/) the MEMS market is expected to grow due to better bio-chemical, bio-medical sensors, and energy harvesters. Beside these last type of devices it is expected, due to a demand on effective farming and a boom of IoT application in agriculture, that the MEMS industry also incorporates sensors developed exclusively to be used in agriculture.
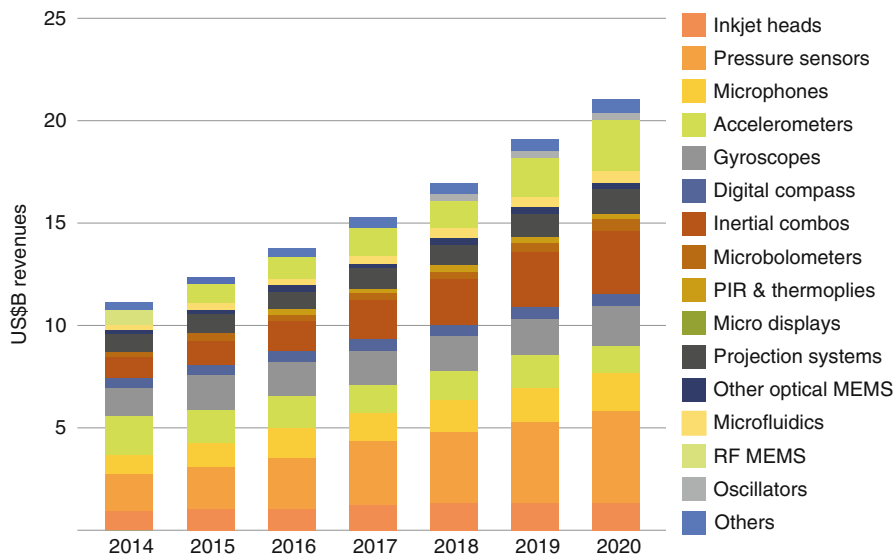
Fig. 17.1 MEMS market forecast: 2014–2020 value (in B$, adapted from Yole Développement, April 2015 [46])

In this chapter it is just considered the MEMS devices primarily developed for agriculture applications—environment, agricultural crops, soil, and livestock.

2   Main Parameters in Agriculture

For clarity the main parameters in Agriculture were grouped into environment and soil, as part of agricultural conditions, and agricultural crops and livestock, as part of agricultural products. All these parameters are summarized in Table 17.1.

2.1   *Environment*

This group of parameters is not an exclusive use in agriculture. Especially, air temperature and humidity are largely used in industry since they are very important for control systems and human comfort. The other addressed parameters are wind direction and velocity, and solar radiation. Although all the spectrum of solar radiation is important for agriculture, only the most important one (photosynthetically Active Radiation—PAR) will be addressed.

Table 17.1 List of main parameters used in agriculture

Type	Classes	Parameters
Agricultural	Environment	Air temperature
Conditions		Air humidity
		Wind direction
		Wind velocity
		Solar radiation
	Soil	Soil moisture content
		Water flux
		Electrical conductivity
		Soil pH
		Soil temperature and thermal properties
Agricultural	Agricultural	Sap flow
Products	Crops	Leaf wetness
		Fruit width
		Stem width
	Livestock	Estrus detection
		Body temperature

2.1.1 Air Temperature and Humidity

Humidity sensors can be classified into categories: capacitive type, which uses a hygroscopic dielectric material between two electrodes forming a capacitor [41]—Fig. 17.2 illustrates an example of this type of sensor; resistive type, where changes in the resistance of sensor element with response to the changes in humidity [59]—Fig. 17.3 illustrates an example of this type of sensor; thermal conductivity type, by taking the difference between the thermal conductivity of dry and wet air (water vapor) [48]; hygrometric type, transduction from air humidity via the mechanical domain [9]; gravimetric type, the change of frequency in thin plates of piezoelectric quartz, coated with a hygroscopic layer, acts as a measure for the humidity [53]; and optical type, e.g. use of the absorption wavelength of water [1]. At the present time, hygroscopic type[1] (capacitive and resistive) humidity sensors are categorized into three classes, including ceramic type (semiconductor), organic polymer-based sensors, and organic/inorganic hybrid sensors (polymer/ceramic) [21]. The most used hygroscopic material for humidity sensors is Al_2O_3 (aluminium oxide), although some other oxide-based sensing material could be used (e.g., TiO_2—Titanium dioxide, SiO_2—silicon dioxide, etc.) [14]. Some of these types of sensors

---

[1]Utilize changes in the physical and electrical properties of the sensitive elements when exposed to the different atmospheric humidity conditions of the surrounding environment, and provide a measure of the humidity due to some amount of adsorption and desorption of water vapor molecules.
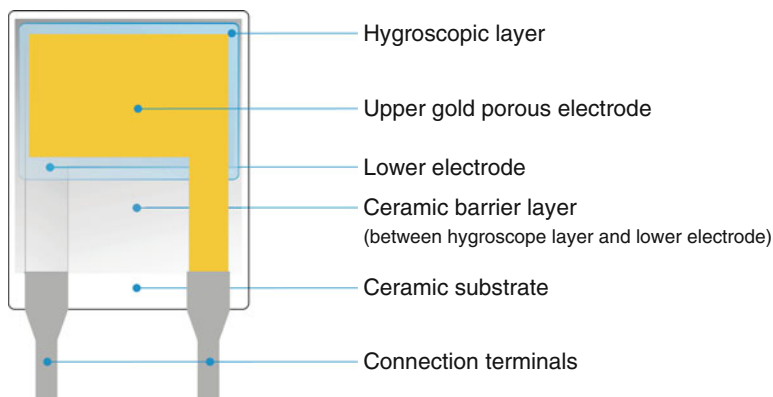
Fig. 17.2 Capacity type humidity sensor
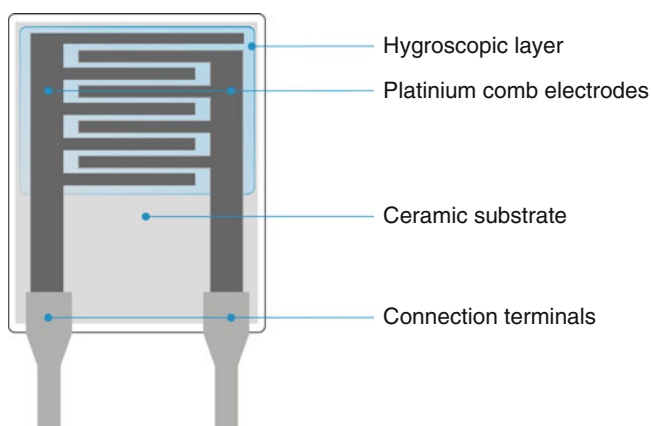
Fig. 17.3 Resistive type humidity sensor

have integrated air temperature sensors, to compensate for temperature dependence or inherent to the method used (case of thermal conductivity type). Therefore, there is no need to use an extra temperature sensor.

MEMS Humidity Sensor

Nowadays, almost all humidity sensors have built-in electronics for calibration, temperature compensation, and have digital communication to read the measured values (humidity and also temperature) and, sometimes, to write parameters. A good review of some miniaturized humidity sensors is presented by Rittersma [57]. Fang et al. [20] recently introduced a new MEMS-based temperature and humidity sensor based thermal conductivity.

2.1.2 Wind Direction and Velocity

To measure wind speed almost all the sensors used in agriculture are mechanical anemometers, usually consisting of propellers or cups to measure the wind speed and vanes to detect the wind direction. Another type of wind speed sensors is the sonic (also named ultrasonic) anemometer which operates by measuring the time taken for a pulse of sound to travel between a pair of transducers. This time depends on the distance between the transducers, the speed of sound, and the air speed along the axis of the transducers. Arranging three pairs of transducers on three different axis, three dimensional wind speed and hence direction and wind angle is obtained [15]. Ultrasonic anemometers represent a valid alternative to mechanical sensors being faster, practically immune to wear and more robust [35], but they are, normally, more expensive.

MEMS Wind Sensor

Due to the advances in MEMS, recently have appeared MEMS-based wind sensors. MEMS-based wind sensor has the advantages of small size, light weight, and low cost. Recently some kinds of wind sensor have been designed based on mechanical principle or thermal principle. Mechanical wind sensor normally uses a cantilever that bends with relation to wind speed [18]. In wind sensors based on thermal principle a heated element is exposed to the wind and the quantity of air velocity is determined by the heat loss. When several heat elements and temperature sensors are arranged in order, both the wind speed and direction can be obtained. The power consumption of thermal anemometers, compared with the other kinds of wind sensors, was a problem solved by MEMS miniaturization and thermal insulation by micro-machining techniques. Zhu et al. [67] have presented an overview and development of 2-D micromachined thermal wind sensors. Thermal wind sensors have two main geometries: in-plane (shown in Fig. 17.4), where the sensor elements are placed in the substrate surface and the airflow is in-plane[68, 69]; out-of-plane (shown in Fig. 17.5), where the sensor elements are placed perpendicular to the substrate surface, normally using a tube, and the airflow is out-of-plane [13, 39, 66]. An out-of-plane airflow sensor has shown greater sensitivity than an in-plane one by elevating the thermal element away from the bottom of the airflow boundary layer and therefore the thermal element is exposed to greater flow speed [13].

2.1.3 Solar Radiation

Solar radiation is a radiant (electromagnetic) energy from the sun. It provides light and heat to the Earth and energy for photosynthesis. This radiant energy is necessary for the metabolism of the environment and their inhabitants. The three relevant bands, or ranges, along the solar radiation spectrum are ultraviolet, visible (PAR), and infrared [24].

Fig. 17.4 In-plane thermal wind sensor by Zhu et al. [68]
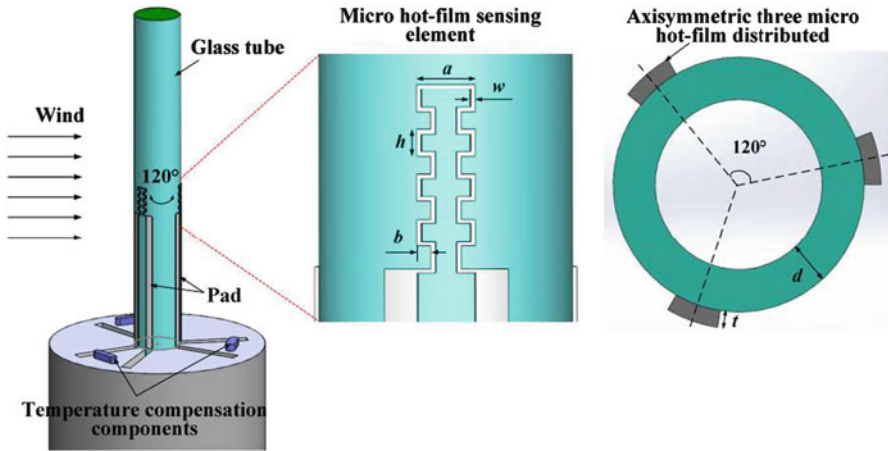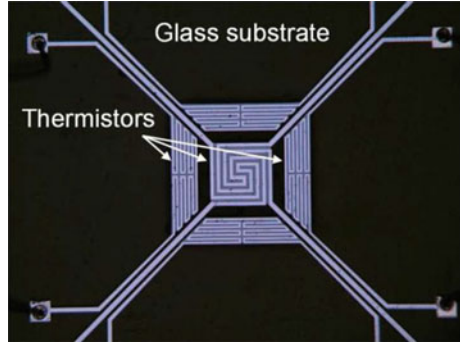
Fig. 17.5 Out-of-plane thermal wind sensor by Yang et al. [66]

Photosynthetically active radiation is the light wavelength range that is best for photosynthesis to occur. Photosynthesis is a process that requires light energy and optimally occurs in the 400–700 nanometer (nm) range. This range is also known as visible light [24].

Several types of photodiode are suitable for light measurement in the visible range. These include silicon (Si), selenium, cadmium sulfide, lead sulfide, lead selenide, and gallium arsenide phosphide (GaAsP) photodiodes. Of these, Si (e.g., S1087 Si photodiode from Hamamatsu [27]) and GaAsP photodiodes are the most useful. A Si photodiode (blue enhanced) is used in the LI-COR quantum sensor (LI-190 SA or SB), which is one of the most widely used for monitoring photosynthetic active radiation [23]. To have an ideal response, as shown in Fig. 17.6, all PAR sensors have an optical filter. Fielder et al. [23] developed a low-cost PAR sensor.
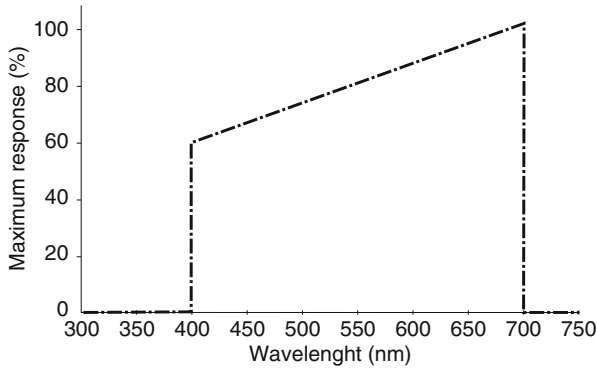
Fig. 17.6 Ideal response of a PAR sensor

MEMS Solar Radiation Sensor

Not considering photodiode or phototransistor a MEMS device and since almost solar radiation sensors used in agriculture have a photodiode/phototransistor and a filter (normally acrylic), there is no MEMS solar radiation sensor for agriculture. It is possible to be made a MEMS device with a Si (or GaAsP) photodiode with a Fabry–Perrot filter (e.g., [31, 32]) in the range of visible light (400–700 nm), or a broader spectrum, to be used in photosynthetic active radiation detection.

2.2 Soil

There are two important soil parameters used in agriculture: soil moisture content and electrical conductivity measurements. Recently, soil pH is becoming a parameter to measure crop productivity and a tool to determine areas where adequate or excessive fertilization has taken place. The other parameters listed in Table 17.1 are commonly included when some of the methods are used by the sensors to read soil moisture content and electrical conductivity. One of the methods used for reading the soil moisture content is single (or dual) probe heat-pulse method (SPHP or DPHP). With the use of these methods (thermal-based methods), in addition to soil moisture it is also possible to obtain: soil temperature, water flux (with DPHP method), soil thermal properties (thermal conductivity, thermal diffusivity, and specific heat capacity) [2, 4–7, 10, 11, 22, 26, 28, 55]. It is also possible with the use of the same probe (needle-like and conductive) to do electrical conductivity readings [33, 44, 45, 65]. The measurement of pH in soil is normally achieved with the use of ion-sensitive field effect transistor (ISFET). ISFET consists of an ion-sensitive electrode (in this case for pH) and a FET. Palaparthy et al. [52] report

some polymers sensitive to pH that can be used on a MEMS device. The change in physical and chemical properties of the polymer material can be converted to stress on the MEMS device.

MEMS Soil Sensors

Although the needed great size of the sphere of influence of sensor readings for soil parameters (compared to the dimensions of some constituents of the soil—stones, grains of sand, clay aggregates, among others), some researchers have developed MEMS sensors to obtain soil moisture. An early example uses CMOS devices built into needles [64] in the order of millimeters, but should not be considered a MEMS device par excellence. Other authors use the sensors or methods used for air humidity and apply them (with different encapsulation—porous medium) to measure soil moisture [34, 50]. In this case they are measuring soil moisture tension (usually measured with a tensiometer[2]) and not soil moisture content. Also, Liu et al. [38] have developed a moisture microsensor based on poly(3,4-ethylenedioxythiophene) polystyrene sulfonate (PEDOT-PSS) conductive polymer, showed in Fig. 17.7. The change in electrical characteristics of the PEDOT-PSS polymer film is used to determine gravimetric water content in soils.

2.3   Agricultural Crops

Automated monitoring of plant responses can offer interesting opportunities for growers to improve climate regulation in order to obtain optimal plant growth and quality [3]. For woody and semi-herbaceous plants, such as trees or fruit vegetables, a range of different plant monitoring devices have been shown to be applicable, e.g. sap flow gauges to measure sap flow [63], displacement transducers to measure stem diameter fluctuations [37, 63] (also applied to fruit width), and detect surface moisture on leaves. Due to the inherent dimensions of the devices used for the

Fig. 17.7 Moisture microsensor based on conductive polymer Liu et al. [38], © 2008, with permission from Elsevier

---
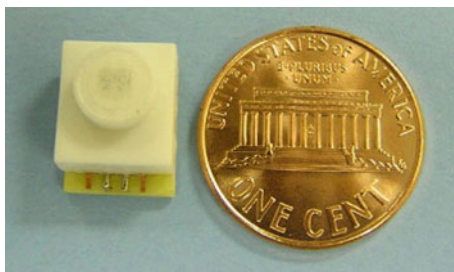
[2]Instrument to measure moisture tension [56].

parameters to be measured, there are few MEMS devices for these purposes. The
parameters they measure are sap flow (or stem water potential) and leaf wetness.

2.3.1 Stem or Fruit Width

The sensor for stem and/or fruit width measures the mechanical displacement
occurred at a cuff placed around the stem (e.g., Ecomatik Diameter Dendrometer
small—DD-S [19], showed in Fig. 17.8) or the fruit (e.g., Dinamax plant growth
sensors [16], as seen in Fig. 17.9). Normally, these devices use linear variable
differential transformer (LVDT) for displacement measurements that leads to stem
or fruit width. Khairi et al. [36] report a device to measure stem width based on
strain gauges.

2.3.2 Sap Flow

The sap flow parameter is important to sense the plant water use. Common methods used for sap flow use heat as a tracer for sap movement, but they are fundamentally different in their operating principles. Two methods are commonly employed, the stem heat balance [58] and trunk sector heat balance [12] methods use the heat balance principle; the stem is heated electrically and the heat balance is solved for the amount of heat taken up by the moving sap stream, which is then used to calculate the mass flow of sap in the stem. In the heat-pulse method, rather than using continuous heating, short pulses of heat are applied and the mass flow of sap is determined from the velocity of the heat pulses moving along the stem. In addition, rates of sap flow can be determined empirically, using the thermal dissipation technique, from the temperature of sapwood near a continuously powered heater implanted in the stem [62].

2.3.3 Leaf Wetness

Leaf wetness sensors are used to detect the presence of surface moisture. Some leaf wetness sensors are used to study the effects of metered vegetative wetting. Others are used to determine the best planting time for crops. Leaf wetness sensors can also be used to monitor the conditions which result in fungal development and growth on plant surfaces. A leaf moisture sensor consists of a chemically treated and gold-plated circuit which changes its electrical resistance in proportion to the moisture which accumulates on the sensing surface [60].

MEMS Leaf Wetness

Using the mentioned methods which use heat or electrical resistance becomes very difficult to use MEMS due to the dimensions involved. Another method that can be used is based on sensing the water potential. This kind of sensor developed by Pagay et al. [50] uses microtensiometer [51] (moisture tension by using a pressure sensor) placed in the leaves to sense wetness or embedded in the stems of woody plants to sense stem water potential (Fig. 17.10).

2.4 Live Stock

For livestock one of the most important parameters is estrus detection, i.e, to monitor and record the event of estrus to perform artificial insemination. For example, detecting estrus and successful impregnating of a cow is an important factor in the

Fig. 17.10 Sensor for measuring water potential by Pagay et al. [50]
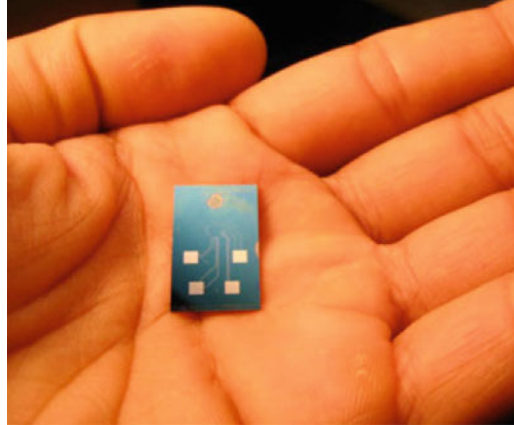
dairy farm [47]. Also, physiological parameters, such as body temperature, are key indicators of heat stress and disease in live stock [30]. Helwatka et al. [29] show a more detailed list of sensors for animal health monitoring.

2.4.1 Estrus Detection

Estrus detection methods most commonly employed by dairy farmers are: visual observation; automatic detection like pedometry, body and milk temperature, and vaginal mucus resistance; chin ball markers; Kamar estrus mount detectors; teaser animals; and progesterone tests. Of course, only the methods that use electronic devices (excluding video observation with digital image processing) are important for this chapter—pedometry, body and milk temperature, and vaginal mucus resistance. Pedometry is made by using accelerometers attached to the cow's leg, and measures the number of steps per hour [61].

MEMS Estrus Detection

The use of MEMS devices is important due to miniaturization required for implantable devices. These type of devices are used for body temperature and vaginal mucus resistance. Morais et al. [43] and following their achievements [42] developed a implantable (vaginal) microsystem device for electrical resistance and temperature measurements in dairy cows, suitable for estrus detection (Fig. 17.11).

2.4.2 Body Temperature

For body temperature in livestock (including poultry) the most common method has been to make spot measurements with a mercury rectal thermometer, and more recently with electronic device, continuous measurements are commonly taken
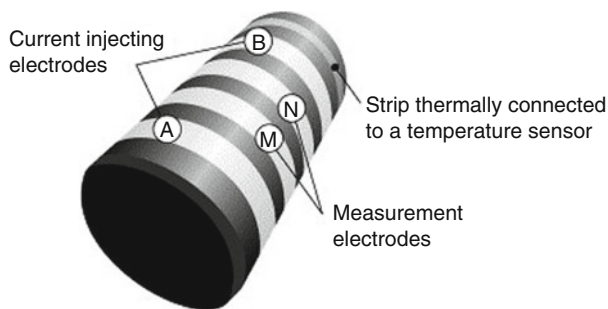
Fig. 17.11 Implantable microsystem device for electrical resistance and temperature measurements. Reprinted from [43], © 2006, with permission from Elsevier

either rectally or tympanically. Rectal probes are easy to insert, and are generally non-invasive. The disadvantage of rectal probes in cattle is that they can only be inserted for a short period of time (5 days) without causing tissue irritation. In poultry, an additional disadvantage is that the birds movement is restricted, and the probe tends to fall out. Tympanic probes can be inserted into the ear canal without anesthetizing the animal by using a headgate or a squeeze chute [8].

MEMS Body Temperature

Normally, the body temperature parameter in MEMS implantable devices used in livestock is part of a multisensor device and not a stand alone parameter. The device presented by Morais et al. [43] for estrus detection also uses body temperature (as part of the measuring process and a stand-alone parameter), and Futagawa et al. [25] have developed an implantable device to sense cellulose content in the rumen. They used electrical conductivity (electrodes placed on a silicon substrate) and a temperature sensor consisted of a p-n junction diode fabricated using complementary metal oxide semiconductor (CMOS) technology (Fig. 17.12).

2.5   Other MEMS Devices

Recently, some new MEMS devices are appearing in the agricultural domain. They don't fit into the classes of Table 17.1 but they are an example of MEMS use in agriculture. One device is a complete weather monitoring system and the other is a device for studying mechanisms of root growth.
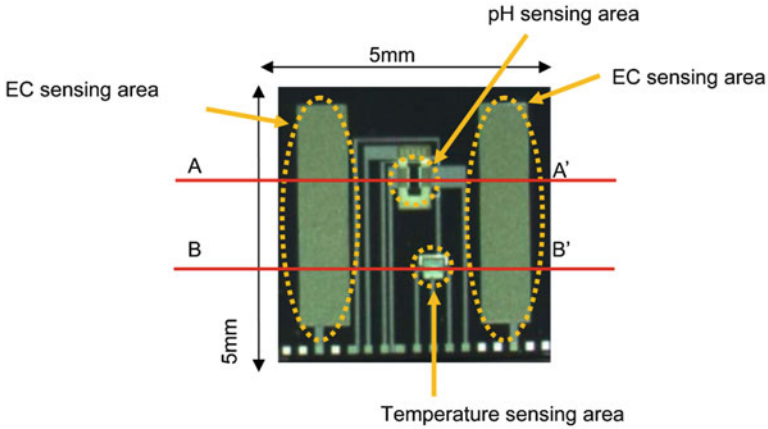
Fig. 17.12 Photograph of the multimodal sensor chip integrated with EC and temperature sensors used in health control of cows [25], © 2011, with permission from John Wiley and Sons
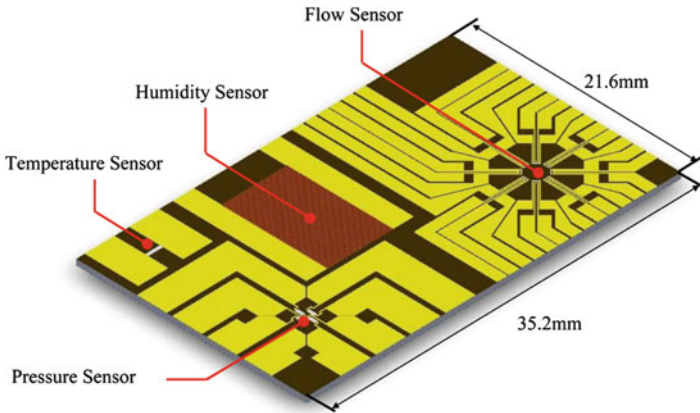
Fig. 17.13 The wireless remote weather monitoring system by Ma et al. [40]

2.5.1 MEMS Weather Monitoring System

Recently, Ma et al. [40] proposed a wireless remote weather monitoring system based on MEMS and wireless sensor network (WSN) technologies comprising sensors for the measurement of temperature, humidity, pressure, wind speed and direction, integrated on a single chip. As shown in Fig. 17.13, the developed sensors comprise a platinum (Pt) resistance temperature detector, a humidity sensor with gold inter-digit electrodes, a Pt-piezoresistor-based pressure sensor, and a flow sensor to determine the direction and velocity of airflow integrated onto a single chip [40].

2.5.2   MEMS to Monitor Root Grow

Ozoe et al. [49] have reported an on-chip analytical method for studying physical mechanisms of plant root growth in soil environments. To quantitatively evaluate physical interaction between root and soil, they developed a silicon-based microchannel device integrated with force displacement sensor which mimics a barrier in soil.

3   Conclusions

In this chapter was presented several MEMS devices where the main application is agriculture. These applications start from humidity sensor, wind speed and velocity, and solar radiation for the environment. Also, in agricultural conditions, applications of MEMS devices to sense soil moisture were showed. For agricultural products like crops and live stock here presented MEMS devices for water potential in stem, to sense moisture in leaves, to detect estrus, and to get livestock body temperature. The range of MEMS devices shown demonstrates that the use of these devices have promising applicability to sense physical parameters commonly used in agriculture, but more work has to be done in order to realize their full potential. Most work with MEMS for agriculture has been done in laboratories thus far. However, transitioning to in situ applications seems feasible. Potential advantages of MEMS are: small size, economical production (specially in large scale), built-in electronics (for auto-calibration, self-testing, digital compensation, and digital communications), and low power consumption (ideal for IoT).

References

1. Bariáin C, Matías IR, Arregui FJ, López-Amo M (2000) Optical fiber humidity sensor based on a tapered fiber coated with agarose gel. Sens Actuators B Chem 69(1–2): 127–131. doi:http://dx.doi.org/10.1016/S0925-4005(00)00524-4. http://www.sciencedirect.com/science/article/pii/S0925400500005244
2. Bilskie J, Horton R, Bristow K (1998) Test of a dual-probe, heat-pulse method for determining thermal properties of porous materials. Soil Sci 163(5):346–355. http://www.soilsci.com/pt/re/soilsci/fulltext.00010694-199805000-00002.htm
3. Bleyaert P, Vermeulen K, Steppe K, Dekock J (2012) Evaluation of a sensor for online measurements of stem diameter variations of leafy vegetables. Acta Hortic 927:571–579. doi:10.17660/actahortic.2012.927.70. http://dx.doi.org/10.17660/ActaHortic.2012.927.70

4. Bristow KL, Campbell GS, Calissendorff K (1993) Test of a heat-pulse probe for measuring changes in soil water content. Soil Sci Soc Am J 57:930–934

5. Bristow K, Kluitenberg G, Horton R (1994) Measurement of soil thermal properties with a dual-probe heat-pulse technique. Soil Sci Soc Am J 58:1288–1294

6. Bristow K, White R, Kluitenberg G (1994) Comparison of single and dual probes for measuring soil thermal properties with transient heating. Aust J Soil Res 32:447–464

7. Bristow K, Kluitenberg G, Goding C, Fitzgerald T (2001) A small multi-needle probe for measuring soil thermal properties, water content and electrical conductivity. Comput Electron Agric 31(3):265–280

8. Brown-Brandl TM, Yanagi T, Xin H, Gates R, Bucklin R, Ross G (2001) Telemetry system for measuring core body temperature in livestock and poultry. In: 2001 Sacramento, CA July 29-August 1, 2001, American Society of Agricultural and Biological Engineers (ASABE). doi:10.13031/2013.23649. http://dx.doi.org/10.13031/2013.23649

9. Buchhold R, Nakladal A, Gerlach G, Neumann P (1998) Design studies on piezoresistive humidity sensors. Sens Actuators B Chem 53(1–2):1–7. doi:http://dx.doi.org/10.1016/S0925-4005(98)00297-4. http://www.sciencedirect.com/science/article/pii/S0925400598002974

10. Campbell GS, Calissendorff K, Williams JH (1991) Probe for measuring soil specific heat using a heat pulse method. Soil Sci Soc Am J 55:291–293

11. Campbell D, Laybourne C, Blair I (2002) Measuring peat moisture content using the dual-probe heat pulse technique. Aust J Soil Res 40(1):177–190

12. Čermák J, Deml M, Penka M (1973) A new method of sap flow rate determination in trees. Biol Plant 15(3):171–178. doi:10.1007/bf02922390. http://dx.doi.org/10.1007/BF02922390

13. Chen J, Liu C (2003) Development and characterization of surface micromachined, out-of-plane hot-wire anemometer. J Microelectromech Syst 12(6):979–988. doi:10.1109/JMEMS.2003.820261

14. Chen Z, Lu C (2005-12-01T00:00:00) Humidity sensors: A review of materials and mechanisms. Sens Lett 3(4):274–295. doi:doi:10.1166/sl.2005.045. http://www.ingentaconnect.com/content/asp/senlet/2005/00000003/00000004/art00002

15. Cuerva A, Sanz-Andrès A (2000) On sonic anemometer measurement theory. J Wind Eng Ind Aerodyn 88(1):25–55. doi:http://dx.doi.org/10.1016/S0167-6105(00)00023-4. http://www.sciencedirect.com/science/article/pii/S0167610500000234

16. Dinamax (2016) Dinamax DEX20, DEX70, DEX100, and DEX200 plant growth sensors. http://www.dynamax.com/products/plant-growth-sensors/dex-fruit-stem-growth-dendrometer, Accessed on 18 Jan 2016

17. Doe P (2015) IoT needs new MEMS approaches. http://www.eetimes.com/author.asp?section_id=36&doc_id=1326768&, Accessed on 18 Jan 2016

18. Du L, Zhao Z, Pang C (2007) Design and fabrication MEMS-based micro solid state cantilever wind speed sensor. In: International conference on information acquisition, 2007. ICIA '07, pp 336–340. doi:10.1109/ICIA.2007.4295754

19. Ecomatik (2016) Ecomatik diameter dendrometer small (DD-S). http://www.ecomatik.de/en/diameter_dendrometer1.php, Accessed on 18 Jan 2016

20. Fang Z, Zhao Z, Wu Y, Zhang B, Wang Y (2004) Integrated temperature and humidity sensor based MEMS. In: Proceedings of international conference on information acquisition, 2004, pp 84–87. doi:10.1109/ICIA.2004.1373325

21. Farahani H, Wagiran R, Hamidon MN (2014) Humidity sensors principle, mechanism, and fabrication technologies: a comprehensive review. Sensors 14(5):7881. doi:10.3390/s140507881. http://www.mdpi.com/1424-8220/14/5/7881

22. Feng C, Liu J, Wei W, Shi Y (2005) Multi-functional soil water velocity instrument based heat pulse probe. In: ICEMI 2005: conference proceedings of the seventh international conference on electronic measurement & instruments, 5(5):580–583

23. Fielder P, Comeau P, of Forests Research Branch BCM (2000) Construction and testing of an inexpensive PAR sensor. Working paper (British Columbia. Ministry of Forests. Research Branch), British Columbia, Ministry of Forests Research Program. http://www.lexingtonwx.com/techdata/Wp53SolarSensor.pdf

24. Fitch K, Kemker C (2014) Solar radiation and photosynethically active radiation. http://www.fondriest.com/environmental-measurements/parameters/weather/solar-radiation/, Accessed on 18 Jan 2016
25. Futagawa M, Ishida M, Ishida M, Sawada K (2011) Study of a wireless multimodal sensing system integrated with an electrical conductivity sensor and a temperature sensor for the health control of cows. IEEE Trans Electric Electron Eng 6(2):93–96. doi:10.1002/tee.20629. http://dx.doi.org/10.1002/tee.20629
26. Ham J, Benson E (2004) On the construction and calibration of dual-probe heat capacity sensors. Soil Sci Soc Am J 68(4):1185–1190
27. Hamamatsu (2014) S1087/s1133 series of Si photodiodes. http://www.hamamatsu.com/resources/pdf/ssd/s1087_etc_kspd1039e.pdf, Accessed on 18 Jan 2016
28. Heitman J, Horton R, Ren T, Ochsner T (2007) An improved approach for measurement of coupled heat and water transfer in soil cells. Soil Sci Soc Am J 71(3):872–880
29. Helwatka A, Riordan D, Walsh J (2014) Sensor technology for animal health monitoring. In: 8th international conference on sensing technology, Liverpool, pp 266–271. http://www.s2is.org/icst-2014/papers/1569961281.pdf
30. Hicks LC, Hicks WS, Bucklin RA, Shearer JK, Bray DR, Soto P, Carvalho V (2001) Comparison of methods of measuring deep body temperatures of dairy cows. In: Stowell RR, Bucklin R, Bottcher RW (eds) Livestock environment VI, Proceedings of the 6th international symposium 2001, American Society of Agricultural and Biological Engineers (ASABE), Louisville, Kentucky. doi:10.13031/2013.7101. http://dx.doi.org/10.13031/2013.7101
31. Hirokubo N, Komatsu H, Hashimoto N, Sonehara M, Sato T (2012) Wideband visible wavelength range MEMS Fabry-Perot tunable filter with calibration system. In: 2012 IEEE Sensors, pp 1–4. doi:10.1109/ICSENS.2012.6411107
32. Hirokubo N, Komatsu H, Hashimoto N, Sonehara M, Sato T (2013) Wideband visible wavelength range MEMS Fabry - Perot tunable filter with highly accurate calibration system. IEEE Sens J 13(8):2930–2936. doi:10.1109/JSEN.2013.2264458
33. Hopmans J, Mori Y, Mortensen AP, Kluitenberg G, Tuli A, Valente A (2006) Multi-functional heat pulse probe measurements of water, heat, and solute transport in the vadose zone. The 18th World Congress of Soil Science
34. Jackson T, Mansfield K, Saafi M, Colman T, Romine P (2008) Measuring soil temperature and moisture using wireless {MEMS} sensors. Measurement 41(4):381–390. doi:http://dx.doi.org/10.1016/j.measurement.2007.02.009. http://www.sciencedirect.com/science/article/pii/S0263224107000255
35. Kato N, Ohkuma T, Kim J, Marukawa H, Niihori Y (1992) Full scale measurements of wind velocity in two urban areas using an ultrasonic anemometer. J Wind Eng Ind Aerodyn 41(1):67–78. doi:http://dx.doi.org/10.1016/0167-6105(92)90394-P. http://www.sciencedirect.com/science/article/pii/016761059290394P
36. Khairi N, Rizam MS, Naimah M, Nooritawati M, Husna Z (2012) Diameter stem changes detection sensor evaluation using different size of strain gauge on dendrobium stem. Procedia Eng 41:1421–1425. International symposium on robotics and intelligent sensors 2012 (IRIS 2012). doi:http://dx.doi.org/10.1016/j.proeng.2012.07.330, http://www.sciencedirect.com/science/article/pii/S1877705812027300
37. Klepper B, Browning VD, Taylor HM (1971) Stem diameter in relation to plant water status. Plant Physiol 48(6):683–685
38. Liu J, Agarwal M, Varahramyan K, IV ESB, Hodo WD (2008) Polymer-based microsensor for soil moisture measurement. Sens Actuators B Chem 129(2):599–604. doi:http://dx.doi.org/10.1016/j.snb.2007.09.017. http://www.sciencedirect.com/science/article/pii/S0925400507007101
39. Liu S, Yang Z, Zhang Y, Xue F, Pan S, Miao J, Norford L (2015) Micro triple-hot-wire anemometer on small sized glass tube fabricated in 5dof uv lithography system. In: 28th IEEE international conference on micro electro mechanical systems (MEMS), 2015, pp 714–717. doi:10.1109/MEMSYS.2015.7051057

40. Ma RH, Wang YH, Lee CY (2011) Wireless remote weather monitoring system based on MEMS technologies. Sensors 11(3):2715–2727. doi:10.3390/s110302715. http://dx.doi.org/10.3390/s110302715

41. Matsuguchi M, Sadaoka Y, Sakai Y, Kuroiwa T, Ito A (1991) A capacitive-type humidity sensor using cross-linked poly (methyl methacrylate) thin films. J Electrochem Soc138(6):1862–1865

42. Miranda N, Morais R, Dias M, Viegas C, Silva F, Serôdio C, Almeida J, Azevedo J, Reis M (2009) Bioimplantable impedance and temperature monitor low power micro-system suitable for estrus detection. Procedia Chem 1(1):505–508. Proceedings of the Eurosensors {XXIII} conference doi:http://dx.doi.org/10.1016/j.proche.2009.07.126. http://www.sciencedirect.com/science/article/pii/S1876619609001272

43. Morais R, Valente A, Almeida JC, Silva AM, Soares S, Reis M, Valentim R, Azevedo J (2006) Concept study of an implantable microsystem for electrical resistance and temperature measurements in dairy cows, suitable for estrus detection. Sens Actuators A Phys 132(1):354–361. The 19th European conference on solid-state transducers. doi:http://dx.doi.org/10.1016/j.sna.2006.04.011. http://www.sciencedirect.com/science/article/pii/S0924424706002949

44. Mori Y, Hopmans JW, Mortensen AP, Kluitenberg GJ (2003) Multi-functional heat pulse probe for the simultaneous measurement of soil water content, solute concentration, and heat transport parameters. Vadose Zone J 2:561–571

45. Mortensen A, Hopmans J, Mori Y, Simunek J (2006) Multi-functional heat pulse probe measurements of coupled vadose zone flow and transport. Adv Water Resour 29(2):250–267

46. Mounier E, Troadec C, Girardin G, de Charentenay Y (2015) Mems markets - status of the MEMS industry 2015. Market & technology report, Yole Développement. http://www.i-micronews.com/images/Flyers/MEMS/Yole_Status_of_the_MEMS_Industry_April_2015_web.pdf

47. Muxlow J (2005) Importance of estrus detection on dairy farms: evaluating pedometers as a possible alternative to visual estrus observation. http://www.afimilk.com/knowledge-center/articles/importance-estrus-detection-dairy-farms-evaluating-pedometers-possible, Accessed on 18 Jan 2016

48. Okcan B, Akin T (2004) A thermal conductivity based humidity sensor in a standard CMOS process. In: 17th IEEE international conference on micro electro mechanical systems, 2004 (MEMS), pp 552–555. doi:10.1109/MEMS.2004.1290644

49. Ozoe K, Hida H, Kanno I, Higashiyama T, Notaguchi M (2015) Early characterization method of plant root adaptability to soil environments. In: 28th IEEE international conference on micro electro mechanical systems (MEMS), 2015, pp 702–705. doi:10.1109/MEMSYS.2015.7051054

50. Pagay V, Lakso A, Strook A (2011) A novel MEMS-based microfluidic water potential sensor for monitoring of water stress in woody plants and soils. In: Shoshany M, Shaviv A (eds) AGRI-SENSING 2011 - international symposium on sensing in agriculture in memory of Dahlia Greidinger, Technion-Israel Institute of Technology, Haifa, p 95. http://agri-sensing.technion.ac.il/ABSTRACT%20BOOK%20final.pdf

51. Pagay V, Santiago M, Sessoms DA, Huber EJ, Vincent O, Pharkya A, Corso TN, Lakso AN, Stroock AD (2014) A microtensiometer capable of measuring water potentials below -10 MPa. Lab Chip 14(15):2806. doi:10.1039/c4lc00342j. http://dx.doi.org/10.1039/c4lc00342j

52. Palaparthy VS, Baghini MS, Singh DN (2013) Review of polymer-based sensors for agriculture-related applications. Emerg Mater Res 2(4):166–180. doi:10.1680/emr.13.00010. http://dx.doi.org/10.1680/emr.13.00010

53. Pascal-Delannoy F, Sorli B, Boyer A (2000) Quartz crystal microbalance (QCM) used as humidity sensor. Sens Actuators A Phys 84(3):285–291. doi:http://dx.doi.org/10.1016/S0924-4247(00)00391-5. http://www.sciencedirect.com/science/article/pii/S0924424700003915

54. Rasher M (2001) The use of GPS and mobile mapping for decision-based precision agriculture. In: Asian GPS conference 2001, last visited on jan-2016

55. Ren T, Noborio K, Horton R (1999) Measuring soil water content, electrical conductivity, and thermal properties with a thermo-time domain reflectometry probe. Soil Sci Soc Am J 63(3):450–457
56. Richards L (1942) Soil moisture tensiometer materials and construction. Soil Sci 53(4):241–248
57. Rittersma Z (2002) Recent achievements in miniaturised humidity sensors - a review of transduction techniques. Sens Actuators A Phys 96(2–3):196–210. doi:10.1016/S0924-4247(01)00788-9. http://www.sciencedirect.com/science/article/pii/ S0924424701007889
58. Sakuratani T (1981) A heat balance method for measuring water flux in the stem of intact plants. J Agric Meteorol 37(1):9–17. doi:10.2480/agrmet.37.9
59. Santha H, Packirisamy M, Stiharu I, Li X, Rinaldi G (2005) A polyimide based resistive humidity sensor. Sens Rev 25(4):271–276
60. Sentelhas PC, Monteiro JE, Gillespie TJ (2004) Electronic leaf wetness duration sensor: why it should be painted. Int J Biometeorol 48(4):202–205
61. Shen W, Chen C, Zheng S, He S, Li M (2015) The design of system about cow activity based on SVM. IJSH 9(3):91–100 doi:10.14257/ijsh.2015.9.3.09. http://dx.doi.org/10.14257/ijsh.2015. 9.3.09
62. Smith D, Allen S (1996) Measurement of sap flow in plant stems. J Exp Bot 47(12):1833–1844. doi:10.1093/jxb/47.12.1833. http://jxb.oxfordjournals.org/content/47/12/1833.abstract, http:// jxb.oxfordjournals.org/content/47/12/1833.full.pdf+html
63. Steppe K, Lemeur R (2004) An experimental system for analysis of the dynamic sap-flow characteristics in young trees: results of a beech tree. Funct Plant Biol 31(1):83. doi:10.1071/ fp03150. http://dx.doi.org/10.1071/FP03150
64. Valente A, Morais R, Couto C, Correia J (2004) Modeling, simulation and testing of a silicon soil moisture sensor based on the dual-probe heat-pulse method. Sens Actuators A Phys 115(2-3):434–439
65. Valente A, Morais R, Tuli A, Hopmans J, Kluitenberg G (2006) Multi-functional probe for small-scale simultaneous measurements of soil thermal properties, water content, and electrical conductivity. Sens Actuators A Phys 132(1):70–77
66. Yang Z, Liu S, Xue F, Zhang Y, Zhao X, Miao J, Norford L (2015) Micro anemometer by a MEMS compatible lab-on-a-tube technology. In: 18th International Conference on solid-state sensors, actuators and microsystems (Transducers), 2015 Transducers - 2015, pp 383–386. doi:10.1109/TRANSDUCERS.2015.7180941
67. Zhu Y, Chen B, Qin M, Huang Q-A (2014) 2-D micromachined thermal wind sensors - a review. IEEE Internet Things J 1(3):216–232. doi:10.1109/JIOT.2014.2319296
68. Zhu Y, Chen B, Qin M, Huang Q-A (2015) A self-packaged self-heated thermal wind sensor with high reliability and low power consumption. In: 2015 IEEE 10th International Conference on nano/micro engineered and molecular systems (NEMS), pp 193–196. doi:10.1109/NEMS. 2015.7147408
69. Zhu Y, Chen B, Qin M, Huang Q-A (2015) Development of a robust 2-d thermal wind sensor using glass reflow process for low power applications. In: 2015 IEEE 65th electronic components and technology conference (ECTC), pp 1633–1639. doi:10.1109/ECTC.2015. 7159815

Chapter 18
MEMS Pressure-Flow-Temperature Sensor for Hydraulic Systems

Charles Groepper, Perry Y. Li, Tianhong Cui, and Kim A. Stelson

Abstract The design, fabrication, and testing of a multifunctional MEMS sensor for use in hydraulic systems is presented. All sensing capabilities are packaged in a single component that can be integrated into a hydraulic system, providing an accurate, low-cost, smaller, and more efficient alternative to individual sensors. The PQT sensor measures pressure, flow, and temperature. Temperature measurement uses a polysilicon thermister. Pressure measurement uses piezoresistive strain elements. Flow is calculated from the differential pressure measurement in an existing flow geometry of the hydraulic system. Design is facilitated by CFD and FEM analysis and the MEMS fabrication and measurement calibration procedure are described in detail. Independent testing of each of the component on the device showed each has good accuracy. The MEMS device was incorporated into a typical fluid power component, an elbow, demonstrating the practicality of the design.

Keywords Integrated sensor • Smart hydraulics • Flow bend sensor • Dimensional analysis • Computational fluid mechanics (CFD)

1 Introduction

Many machines use hydraulics as a means of energy transmission to take advantage of hydraulics' capability to deliver high power and high force or torque in small packages. Some examples include metal presses, injection molding machines, assembly line robots, construction and agricultural equipment, off-road vehicles, and cranes. To control hydraulic machines accurately and to monitor machine health, real time measurements of pressure (P), flow (Q), and temperature (T) are

C. Groepper
AT Instruments, Cardiff, UK

University of Minnesota, Minneapolis, MN, USA
e-mail: groepper@hotmail.com

P.Y. Li (✉) • T. Cui • K.A. Stelson
University of Minnesota, Minneapolis, MN, USA
e-mail: lixx009@umn.edu; cuixx006@umn.edu; kstelson@umn.edu

needed. Pressure determines force or torque, flow determines speed and temperature determines fluid viscosity and indicates system efficiency. Measuring these variables is especially important for the emerging area of smart hydraulics where intelligent functions are achieved using sensing, control, and diagnostics. Traditionally, the measurement of P, Q, and T is accomplished by adding discrete sensing components to the fluid power system. This increases weight, size, and cost of the system. In this paper, an alternative approach which integrates small, multifunction MEMS sensors into existing fluid power components to measure the pressure, flow, and temperature is described. While pressure and temperature sensing using MEMS devices is not new, an innovation is the use of pressure information in the existing flow geometry of the application to measure flow. In other flow measurement approaches, additional devices are added to the flow path inevitably changing the flow and increasing pressure drop and inefficiency. For example, an orifice meter adds an orifice to the flow path and measures flow from the pressure drop; a turbine meter adds a turbine to the flow path and measures flow from the turbine speed; a Coriolis flow meter adds a vibrating flow bend and measures flow from the oscillating motion of the bend. The approach here does not create additional pressure drops and is hence more energy efficient. Since the pressure-flow relation depends on fluid viscosity and density and both are functions of temperature, the effect of viscosity and density variation due to temperature changes can be compensated by incorporating temperature sensing into the system. By integrating the sensor into the hydraulic system, a lower cost and more compact solution is created. The sensor is also robust since it does not rely on moving parts.

MEMS pressure and temperature sensing is a mature technology [1–5] and have found many applications. For example, ultra-high sensitive pressure sensors are used as tactile sensors for a robot hand in [6]; by combining with piezoelectric technology, pressure sensors can become self-powered [7] and implantable in medical equipment [8]. Low power and energy-efficient CMOS temperature sensors have been used for temperature compensation in MEMS frequency references [9] and for Radio Frequency Identification (RFID) tags [10]. Fully printed high-sensitivity multifunctional artificial whisker sensors have also been developed by integrating strain and temperature sensors using printable nanocomposite ink [11].

In contrast, the development of MEMS based flow rate sensors for the macro-domain is lacking. Currently, the primary thrust of MEMS flow measurement research is for microfluidic systems and is not appropriate for macro scale flow rate in typical hydraulic systems. Examples include graphene and carbon nanotube based approaches in [12, 13] and the integrated sensor for detecting flow rate as well as gas and liquid compositions in [14]. Other flow sensors use thermal and magnetic effects that are very restrictive to the choice of fluid and are inappropriate for typical hydraulic applications [15].

The primary challenge with the proposed flow sensing method is the establishment of the pressure-flow correlations in the existing geometry of the application. In this work, we illustrate our sensor concept by utilizing a flow bend (or elbow) for flow measurement due to their presence in nearly all systems. Other geometries that can be used include sudden expansions and contractions that also have a relationship

between pressures at various locations and flow. In the literature, the use of elbows to measure flow dates from the early 1920s focuses mainly on the use of large elbows for measuring water flow in irrigation systems where turbulent flow conditions are expected [16–28]. In typical hydraulic applications, the flow is laminar, and further research is needed to evaluate the use of an elbow for flow measurement in laminar flow. Since the design will infer flow rate from the measurement of small pressure differences, a reliable calibration scheme for the pressure-flow relationship is essential to achieve good accuracy.

In this work, piezoresistive strain type MEMS pressure sensors will be used for sensing absolute pressure and differential pressures (for measuring flow). Such sensors have previously been developed and are commercially available for both large and small pressure ranges (e.g., 0–5000 Pa in [1] and 0–50 MPa in [29]). For temperature sensing, thermally sensitive semiconductor resistors will be used. Previous work done by Steinhart and Hart [3] has characterized the temperature behavior of semiconductor resistors using a simple logarithmic polynomial curve fit calibration equation. Following this approach will allow temperature measurement to be easily integrated into the sensor design.

The integrated pressure, flow, and temperature (PQT) sensing method is shown in Fig. 18.1. Development of the sensor follows these steps:

1. validate the flow rate measurement principle through the use of analysis, computation fluid dynamics (CFD) and experiments;
2. develop a calibration procedure that takes into account the variation of the fluid viscosity and density with temperature.
3. design and fabricate a single die multifunction MEMS based sensor
4. integrate the MEMS sensor into an assumed flow geometry and characterize the resulting PQT sensor.
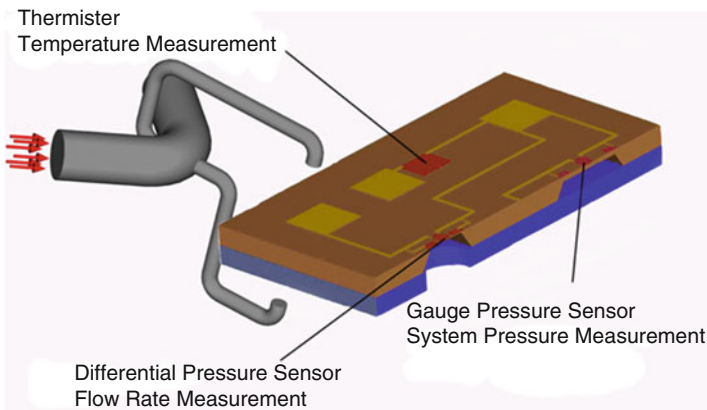
Fig. 18.1 A pressure-flow-temperature sensor integrated into a component

The remainder of this chapter is organized as follows. In Sect. 2, the use of a flow elbow for flow measurement is confirmed by analysis, numerical computation, and experiments. Section 3 presents two calibration approaches. The design and fabrication of the MEMS sensor is presented in Sect. 4. Performance evaluation of the MEMS sensor integrated into a flow elbow is reported in Sect. 5. Concluding remarks are given in Sect. 6.

2 Using Pressure Difference Across an Elbow to Measure Flow

Most hydraulic systems include several flow bends in the flow path. In this section, we illustrate how a 90° flow bend (i.e., an elbow), such as the one shown in Fig. 18.2 can be used for flow rate measurement. First, the Euler equation based on simplifying flow assumptions is presented in Sect. 2.1. This is followed by computational fluid dynamics (CFD) analysis which allows for investigation of important design and flow features.

2.1 Analytical Relation

Consider the 2D center cross-section of the elbow with radius of curvature S and pipe radius r as shown in Fig. 18.2. Assume that the flow is inviscid and is uniform across the radial cross section. Applying the momentum equation for a differential fluid element in the direction normal to the streamline results in the Euler equation for the pressure $p(n)$ [30]:

$$\frac{1}{\rho}\frac{\partial p(n)}{\partial n} = \frac{v^2}{n} \tag{18.1}$$

Fig. 18.2 A cross-section of an elbow flow conduit
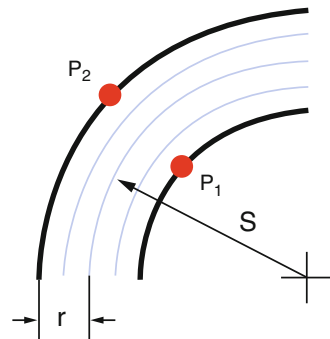
where n is the distance of the fluid element from the center of curvature and v is the fluid velocity. Assuming that the flow velocity v is uniform over the cross-section of the elbow, the flow rate is $Q = \pi r^2 v$. Equation (18.2) can be integrated from the inner radius ($n = S - r$) to the outer radius ($n = S + r$) to obtain:

$$Q = \pi r^2 \sqrt{\frac{P_2 - P_1}{\rho \ln\left[\frac{S+r}{S-r}\right]}} \tag{18.2}$$

Despite the simplifying assumptions of inviscid flow and a constant velocity profile across the pipe, Eq. (18.2) shows a clear dependence of the pressure difference between the inner and outer radii of the elbow on flow rate. This suggests that the pressure difference between the inner and outer radii can be used to measure flow rate.

2.2   CFD Analysis

In this section, we present a CFD analysis of the flow in an elbow. In addition to pressure profile versus flow relationships, effects of entrance and exit lengths, fluid viscosity and secondary flow can be studied. In this study, mapped brick meshes are used. This meshing method is known to offer fast solution times and control over element placement. The meshes have also been refined sufficiently to ensure accurate results. Approximately, 120,000 nodes are used. Computation is done using the ANSYS CFX software. The hydraulic fluid is Mobil DTE 25.

2.2.1   Pressure-Flow Relation

A CFD computation was conducted on flow elbows with nominal sizes of 1/8," 1/4," and 3/8." Their respective inner pipe radii r, are 5.2, 6.5, and 8.35 mm. Their respective radii of curvature S are 7.3, 10.4, and 11.2 mm. Figure 18.3 shows the comparison between the predictions from CFD and the simplified Euler equation (18.2) for the 1/8" elbow (results for the other two elbows are similar). Note that the maximum flow is 15.3 LPM. Figure 18.3 shows that despite the simplifying assumption of the Euler equation, its prediction is quite close to that of the CFD. Experimental results in Fig. 18.3 were obtained from a standard commercial cast iron elbow. The threads were removed to ensure a smooth flow surface and pressure taps were drilled in the elbow using the smallest drill bit (1.6 mm) that can be used without breakage. Both the CFD and Euler equation predict pressure differences that are smaller than seen in the experiments. However, the strong trend of increasing pressure difference as flow increases is apparent in both experiments, the CFD computation and the Euler equation. Error exists in estimating the pipe diameter. Also, the actual elbow includes fittings immediately before and after the elbow that

Pressure Difference as a function of Pipe Velocity for 1/8 '' Elbow
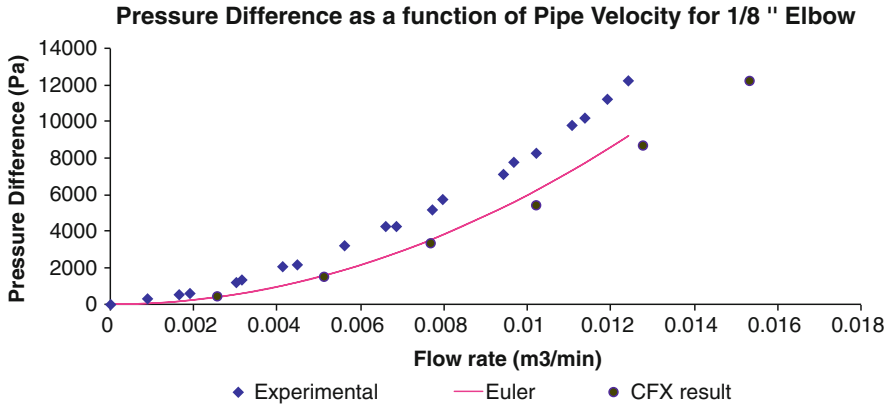
Fig. 18.3 Comparison between Euler equation, CFD and experimental pressure versus flow rate results for a 1/8" elbow

Fig. 18.4 A cross section of the commercial elbow used in the experiments represented in Fig. 18.3

SECTION   A-A
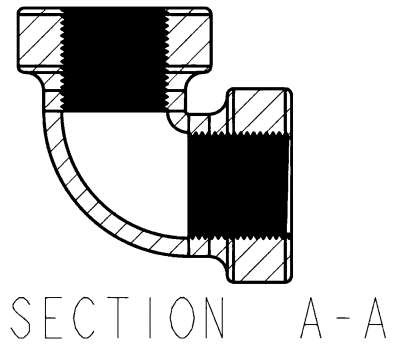
induce sudden expansion and contraction. These are not modeled in the simple analysis or the CFD which have assumed sufficient entry and exit lengths for fully developed flow to occur. This difference and other geometry differences highlight the importance of a robust calibration scheme that uses experimental data (Fig. 18.4).

2.2.2 Entrance Length

It is shown in [31] that the entrance length (in pipe diameters) required to establish fully developed flow at the elbow for the laminar flow regime is $0.06R_e$ where R_e is the Reynolds number. For a typical mid-range Reynolds number of 650, an entrance length of 40 pipe diameters is necessary. Such entrance lengths are usually not present and therefore, fully developed flow should not be expected in the bend in
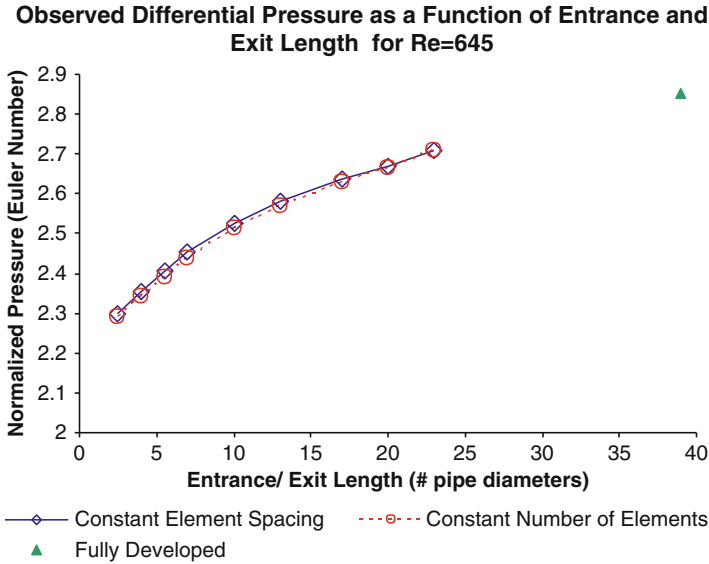
Fig. 18.5 Differential pressure between inner and outer radii normalized by $0.5\rho v^2$ at various entrance lengths. Reynold number of 650 is assumed

most applications. CFD analysis is performed to determine the effect of reducing the entrance length. For each entrance length, a constant velocity profile (a worst case scenario) is assumed at the entrance. Figure 18.5 shows that reducing entrance length reduces the differential pressure between the inner and outer radii of the elbow. If the entrance length is only 2.5 pipe diameters, the differential pressure would be 20 % lower than that of a fully developed case.

2.2.3 Pressure/Velocity Distributions and Secondary Flow

Fluid viscosity affects the flow through a curved pipe by slowing down the fluid, thus creating a non-uniform velocity profile (boundary layer). Also, momentum effects create the pressure difference between the inner and outer radii of the bend. The fluid in the stagnant boundary layer is subject to this pressure difference, and a secondary flow is induced. The study of the magnitude and development of the secondary flow is important to determine the optimal location for the pressure taps in the elbow. Placement of the taps in a recirculation zone or otherwise disturbed region will result in erratic and unreliable behavior.

Figure 18.6 shows the baseline pressure and velocity distributions on the cross sectional (symmetry) plane of the elbow. Here, a fully developed inlet velocity profile, with a Reynolds number of 550 is used. Notice that the velocity is clearly not uniform in the elbow as assumed in Sect. 2.1. However, it is also clear that there is a distinct pressure differential between the inner and outer radii.
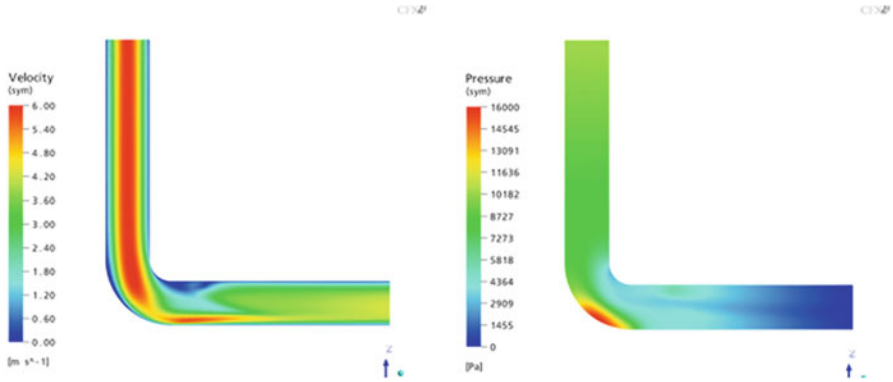
Fig. 18.6 Velocity distribution (*left*) and pressure distributions (*right*). Reynolds number is 550 and the inlet velocity profile is uniform

Fig. 18.7 Definition of the cross-planes
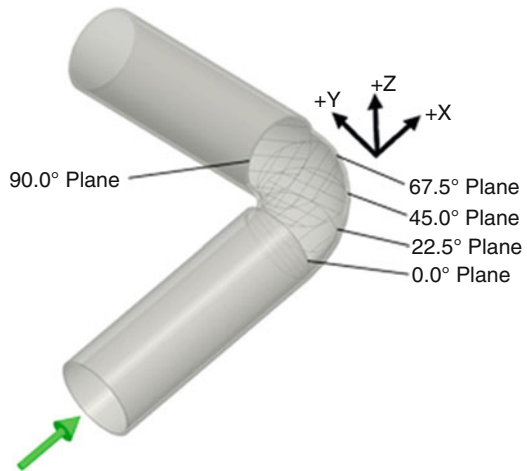
The *z*-direction (upward) velocity distributions at different cross-planes normal to the flow (as defined in Fig. 18.7) are shown in Fig. 18.8. The non-uniform flow with different signs indicate a circulating secondary flow. Note that the magnitude of the secondary flow increases towards the middle and latter portion of the elbow (increasing cross-plane angles). For example, at the 67.5 ° cross-plane, the magnitude of the z component of the velocity reaches 20 % of the average bulk flow velocity.

Another way to view the secondary flow fields occurring in the elbow is through the use of vector plots. With a vector plot, the velocity of each node in the mesh is represented by the magnitude and direction an arrow. A vector plot showing the
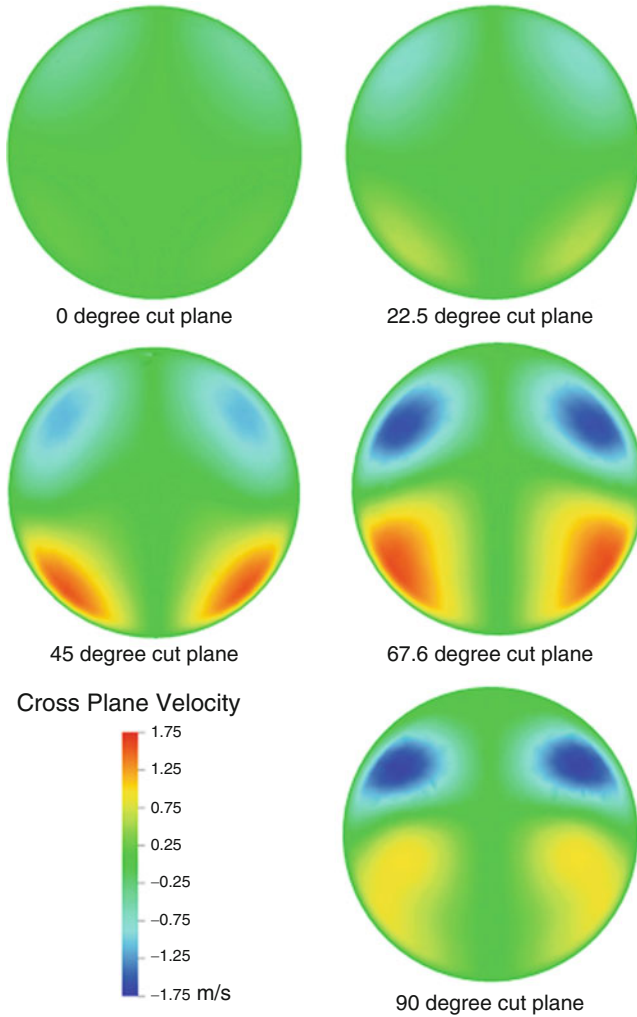
Fig. 18.8 Secondary flow characteristics as indicated by velocities in the *z*-direction at various cross planes. Here, the inside and outside of the elbow, are respectively, at the bottom and top of each circular section. The *z*-direction is horizontal from *right* to *left*

velocity on the 45° cut plane is shown in Fig. 18.9. The circulating secondary flow can be clearly seen.

2.2.4 Pressure Tap Locations

The analysis so far has assumed that the pressure taps are located on the 45° cross-plane. This location does not provide the maximum differential pressure signal, but
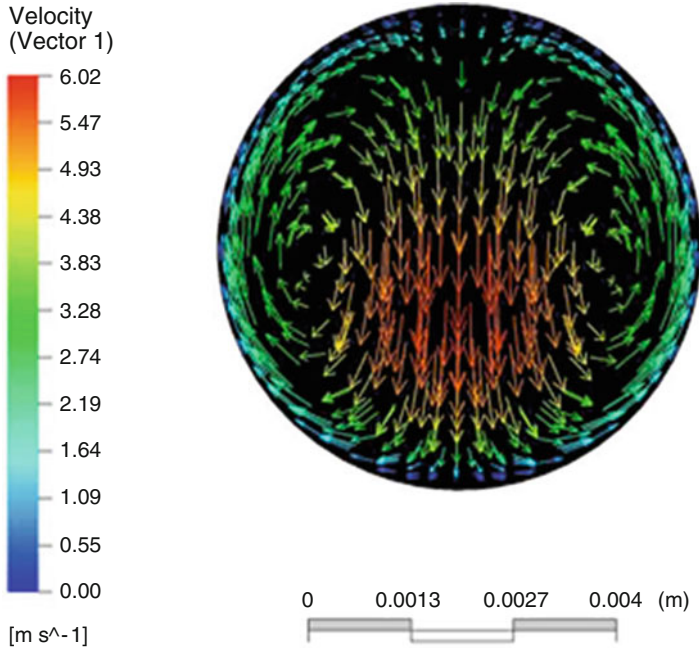
Fig. 18.9 Vector plot of the secondary flow field at 45°

does allow the elbow to be used as a bi-directional meter. If bi-directional service is not needed, more suitable pressure tap locations offering a higher differential output signal can be found. To perform the investigation, CFD was used to examine the pressure distributions along the walls of the elbow. From Fig. 18.10, it is evident that the location of the outer pressure tap provides the greatest influence on the differential pressure output signal. The optimum location for the outer tap is 55–60° with respect to the inlet. For the inner tap, the strongest pressure signal is obtained at approximately 80°. However, this is not a suitable location, as the flow separates from the wall at an angle of 45–50°. A more suitable location for the inner pressure tap is at the local minimum which occurs at an angle of approximately 25–30° from the flow inlet.

The finite diameter of the pressure tap introduces an error that is proportional to the diameter and length of the tap and to the wall shear at the pressure tap locations [32]. For this reason, wall shear should also be evaluated to determine suitable pressure tap locations. Figure 18.11 shows the profiles of the wall shear magnitude for an elbow with fluid entering from the top. The standard 45° plane is in an area of low wall shear, as are the pressure tap locations of 25° on the inner wall and 55° on the outer wall as suggested above to increase pressure signals. The worst locations for pressure taps from a wall shear standpoint would be on the outer radius immediately following the elbow, and on the inner radius immediately preceding the elbow.
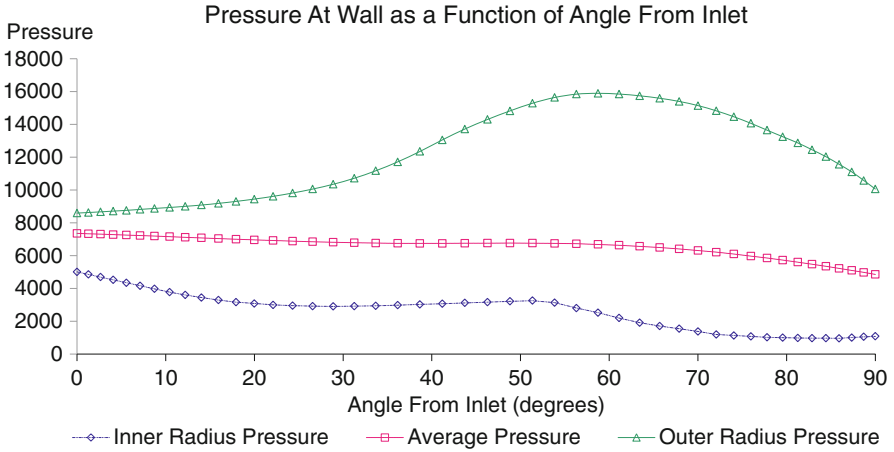
Fig. 18.10 CFD computed pressure distributions on the outer and inner walls of a typical elbow geometry

Fig. 18.11 Contour map showing wall shear magnitude
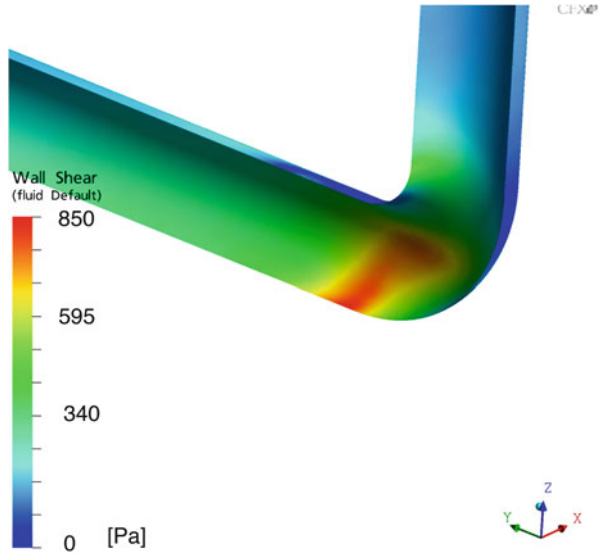
3 Elbow Flow Sensor Calibration

In this section, we develop a calibration process for the elbow flow sensor. Although the method is developed for an elbow, it can also be applied to other geometries that induce repeatable relationships between pressure difference and flow. The calibration process must take into account temperature dependence, as temperature affects the fluid viscosity and density.
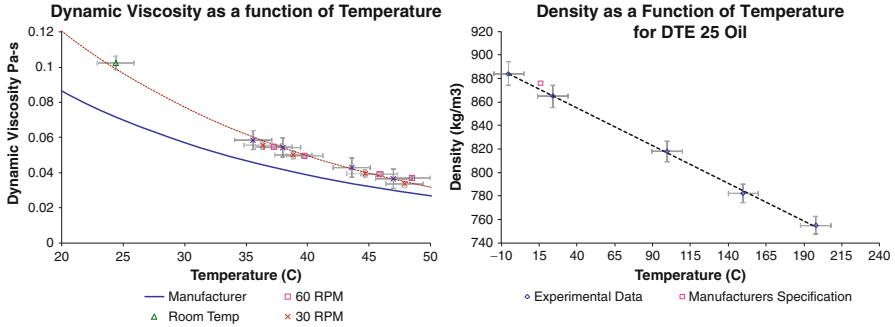
Fig. 18.12 Experimentally measured viscosity-temperature (*left*) and density-temperature relationships of Mobile 1 DTE 25 hydraulic oil

First, the density-temperature and viscosity-temperature relationships of the hydraulic fluid are obtained. For example, the experimentally obtained relationships for the DTE 25 hydraulic oil are shown in Fig. 18.12. Data provided by the manufacturers can also be used if available and accurate. From Fig. 18.12, it is seen that the fluid decreases viscosity by three-fold with only a 20 °C increase in temperature, whereas density decreases nearly linearly with temperature at a rate of approximately 2.5 % per 20 °C.

3.1 Dimensional Analysis

The calibration schemes use the Buckingham Pi dimensional analysis. This approach allows for compensation of all pertinent variables, which is necessary to account for temperature effects.

The basic theory of dimensional analysis states that [33]:

> If an equation involving k variables is dimensionally homogeneous, it can be reduced to a relationship among $k - r$ independent dimensionless products, where r is the minimum number of reference dimensions required to describe the variables.

For the case of laminar flow in an elbow, there are 3 reference dimensions (force-F, length-L, time-T) and 5 variables: (1) pressure difference ΔP, (2) density—ρ, (3) dynamic viscosity μ, (4) flow velocity v, and (5) diameter $D(= 2r)$, resulting in $5 - 3 = 2$ dimensionless variables. Here, we have made the simplifying assumption that the radius of curvature and pipe diameter is proportional to each other. One possible set of dimensionless variables is to have dependence on ΔP only in one variable (π_1), dependence on v only in the other (π_2), and dependence on ρ, μ, and D in either π_1 or π_2. With this in mind,

$$\pi_1 := \frac{\Delta P \rho D^2}{\mu^2} = \left(\frac{\rho v D}{\mu} \right)^2 \cdot \left(\frac{\Delta P}{\rho v^2} \right) = R_e^2 \cdot \text{Eu} \qquad (18.3)$$
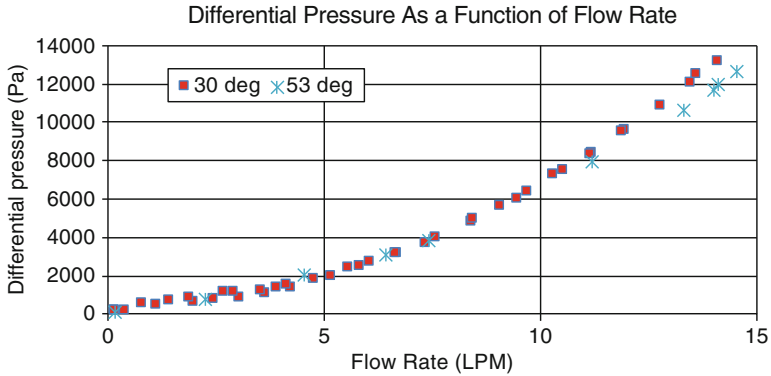
Fig. 18.13 Raw ΔP versus flow rate Q for a 3/8" elbow at two temperatures
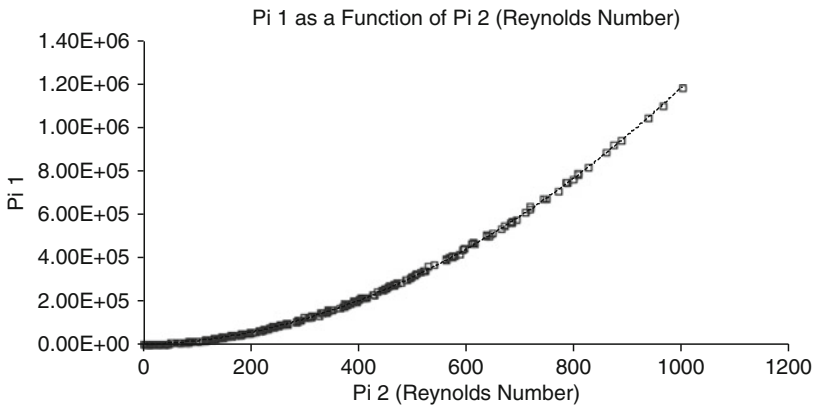
Fig. 18.14 Calibration curve using π_1 and R_e as variables

$$\pi_2 := \frac{\rho v D}{\mu} = R_e \tag{18.4}$$

where $R_e = \frac{\rho v D}{\mu}$ is the Reynolds number and $\mathrm{Eu} := \frac{\Delta P}{\rho v^2}$ is the Euler number. Notice that π_1 does not depend on velocity v as intended, although R_e and Eu do depend on v individually.

The dimensional analysis is applied to the ΔP versus Q data collected at different temperatures such as in Fig. 18.13 where the dependence on temperature is clearly seen. The relation between π_1 and π_2 (which is also the Reynolds number) is shown in Fig. 18.14. Note that the temperature dependence in Fig. 18.13 has been successfully collapsed into a single curve.

The relationship between π_1 and R_e in Fig. 18.14 can be characterized by a second order polynomial curve:

$$\pi_1 = C_1 R_e^2 + C_2 R_e \tag{18.5}$$
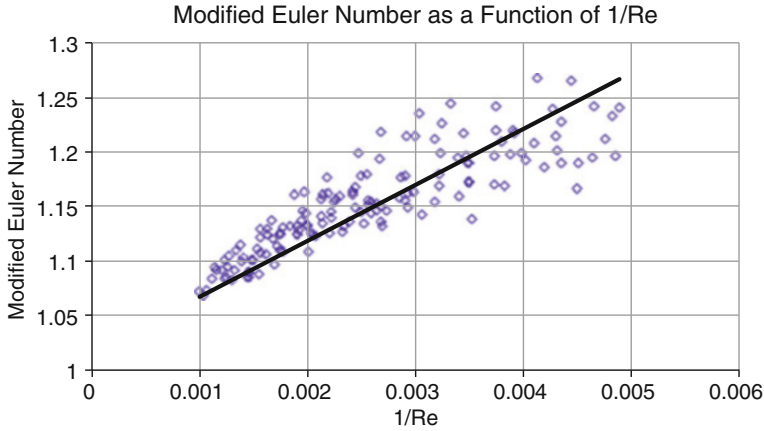
Fig. 18.15 Calibration curve using Eu_{mod} and R_e as variables

where C_1 and C_2 are obtained using least squares method. Rearranging and using the relationship $v = 4Q/(\pi D^2)$ results in the quadratic equation:

$$\Delta P = \frac{16 C_1 \rho(T)}{\pi^2 D^4} Q^2 + \frac{4 C_2 \mu(T)}{\pi D^3} Q \tag{18.6}$$

where density ρ and dynamic viscosity μ are temperature dependent. Given ΔP, the flow rate Q can be solved from (18.6) using the quadratic formula.

An alternative dimensionless variable to π_1 in Eq. (18.3) is suggested by the inviscid, 2D analytical Euler equation (18.2). Define the modified Euler number as:

$$\pi_3 = Eu_{mod} := \frac{\Delta P}{\rho v^2 \ln\left[\frac{R+r}{R-r}\right]} \tag{18.7}$$

where R is the radius of curvature and r is the radius of the pipe. The analytical Euler equation (18.2) would predict that $Eu_{mod} = 1$.

Substituting the experimental results in Fig. 18.13 into Eu_{mod} and R_e in (18.7) and (18.4) gives the relation as shown in Fig. 18.15. Reciprocal of R_e is plotted instead of R_e to emphasize the Euler solution when the fluid is inviscid (i.e., $\mu = 0 \rightarrow 1/R_e = 0$). Although significant scatter is present, a distinct linear relation is seen. In particular, as $1/R_e \rightarrow 0$, Eu_{mod} does extrapolate to 1 as predicted. Thus, this curve is parameterized as:

$$Eu_{mod} = 1 + \frac{C_3}{Re} \tag{18.8}$$

where C_3 is obtained from linear least squares. Substituting (18.4) into (18.8), we have

$$\frac{\Delta P}{\rho v^2 \ln\left[\frac{R+r}{R-r}\right]} = 1 + C_3 \left[\frac{\mu}{\rho v D}\right] \tag{18.9}$$

Rearranging and using the relationship $v = Q/(\pi r^2)$ results in the quadratic equation:

$$\Delta P = \frac{\rho(T)}{\pi^2 r^4} \ln\left[\frac{R+r}{R-r}\right] Q^2 + \left(\frac{C_3 \mu(T)}{2\pi r^3}\right) \ln\left[\frac{R+r}{R-r}\right] Q \tag{18.10}$$

Thus, given ΔP, Q can be obtained by solving the quadratic equation (18.10). One advantage of the method using $\mathrm{Eu_{mod}}$ and R_e is that only one calibration parameter, C_3, is needed, compared to two parameters, C_1 and C_2, if π_1 and R_e are used.

3.2 Calibrated Elbow Meter Performance

To evaluate the performance of the calibration process, a precision machined elbow was manufactured and the flow sensing methodology was applied and evaluated. The elbow was CNC machined from ANSI 1020 steel. It contains two flow channels, one with an internal diameter of 1/4" (6.35 mm) and the other with an internal diameter of 3/8" (9.525 mm). Both channels were machined with a radius to diameter ratio of unity, to allow for direct dimensionless comparisons. The tolerance on the diameter is ± 0.001" (± 0.025 mm). The elbow was designed such that the entrance length could easily be adjusted using standard off-the-shelf pipe nipples. For a nipple to be used in the apparatus, the only requirement was a reaming procedure to match the internal diameters to that of the billet elbow. One disadvantage to this method was the introduction of a sudden expansion/contraction region only a few pipe diameters from the curvature. However, extensive CFD analysis showed the addition of a sudden expansion/contraction region would have only a marginal effect on the measured pressure differential. A three-dimensional rendering of the experimental elbow is shown in Fig. 18.16.

The flow rate was measured with a JVA-60KG positive displacement gear meter (AW company), differential pressure was measured with a Senotec 0–5 psi pressure gage, system pressure was measured with a BarksDale 425 pressure sensor, and temperature was measured with K-type thermocouples.

The elbow meter was characterized in terms of its linearity, repeatability, and average accuracy. With the exception of accuracy, they are reported as a percentage of the full scale capability of the meter. For the 1/4" (6.35 mm) elbow, the full scale capability is assumed to be 12.5 LPM, while the 3/8" (9.525 mm) channel is assumed to have a full scale capability of 15 LPM. Linearity is defined as the maximum deviation from a straight line fit to the data. Performance of the elbow

Fig. 18.16 Render of the precision machined billet elbow
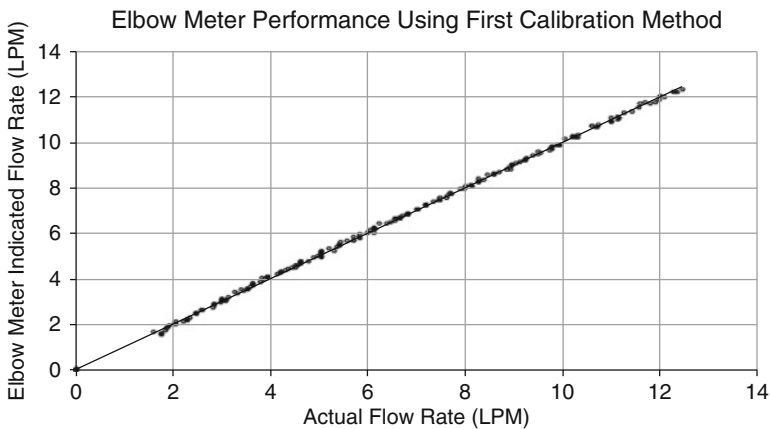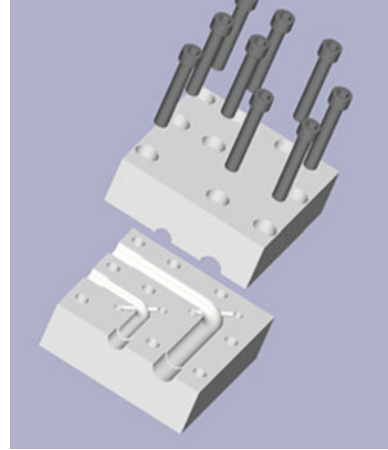
Fig. 18.17 Elbow meter performance curve using π_1 and R_e

meter was evaluated by showing its flow measurement as a function of the actual flow rate. Repeatability was calculated based on the standard error computed as: $\sigma = \sqrt{\sum_{i=1}^{n} e^2/(n-2)}$ where e_i is the error of the $i-th$ sample and n is the number of samples.

When π_1 and R_e variables are used for calibration, for the 1/4" (6.35 mm) elbow meter, the performance curve is shown in Fig. 18.17. It exhibits a linearity of 1.7 % and an average accuracy of 1.5 % over a turndown range of 10:1. The linearity and average accuracy are improved to 1.5 and 1.3 %, respectively, if the turndown of the meter is restricted to 8.0:1. For the 3/8" (9.525 mm) elbow meter, it exhibits a linearity of 6.6 % and an average accuracy error of 7.4 % over a Reynolds number range of 25–350. If the turndown of this meter is restricted to 8.0:1, the linearity is 6.6 % and the average accuracy improves to 6.1 %.
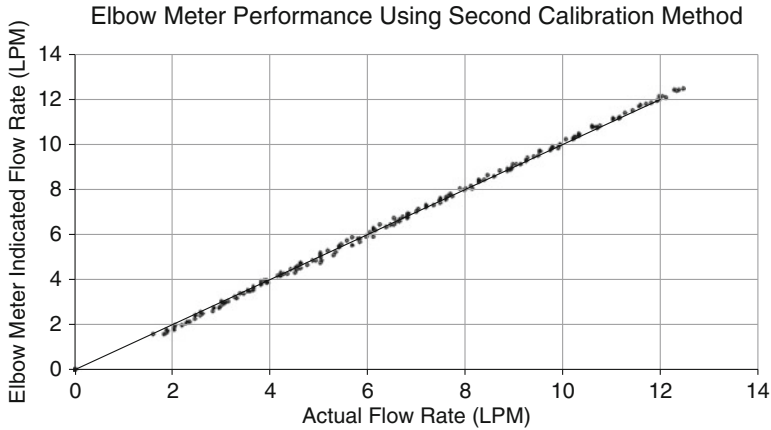
Fig. 18.18 Elbow meter performance curve using Eu_{mod} and R_e

When Eu_{mod} and R_e variables are used for calibration instead, for the 1/4"
(6.35 mm) elbow meter, the performance curve is shown in Fig. 18.18. It exhibits
a linearity of 3.0 % and an average accuracy of 2.0 % over a turndown ratio of
10:1. If the turndown is restricted to 8.0:1, the performance of the meter improves
with a linearity of 2.3 % and average accuracy of 1.9 %. For the 3/8 " (9.525 mm)
elbow meter, it exhibits a linearity of 8.3 % and an average accuracy of 9.1 % with a
turndown ratio of 10:1 when calibrated with the second method. If the turndown
is restricted to 8.0:1, the linearity improves to 7.5 % and the average accuracy
improves to 8.4 %.

These results are summarized in Table 18.1. It is apparent that using π_1 and
R_e as calibration variables exhibits better linearity and accuracy. This is especially
true for low flow rate as is evident in Fig. 18.18. It is also apparent that accuracies
decrease with the larger diameter elbow. This is due to the lower velocity and lower
differential pressure (signal level) for the same flow rate. This suggests that one
should choose a smaller elbow in the application for flow monitoring if available.

Although the performance of the elbow meter is diminished when Eu_{mod} and R_e
are used as calibration variables, the method depends less on the fluid viscosity.
Therefore, the method is expected to exhibit superior performance results with
greater uncertainty in the viscosity.

4   Fabrication

In this section, the design and fabrication of the integrated MEMS pressure, flow
and temperature (PQT) sensor will be presented. Since flow is measured from
differential pressure signals arising from fluid flow through the elbow geometry,

Table 18.1 Elbow flow meter performance

Elbow size	1/4" diameter				3/8" diameter			
Turndown ratio	10:1		8:1		10:1		8:1	
Calibration method	π_1, R_e (%)	Eu_{mod}, R_e (%)	π_1, R_e (%)	Eu_{mod}, R_e (%)	π_1, R_e (%)	Eu_{mod}, R_e (%)	π_1, R_e (%)	Eu_{mod}, R_e (%)
Linearity	1.7	3.0	1.5	2.3	6.6	8.3	6.6	7.5
Average accuracy	1.5	2.0	1.3	1.9	7.4	9.1	6.1	8.4
Repeatability	1.13	1.42	1.20	1.37	2.78	2.39	2.68	2.31

the MEMS device must have the capability to measure gauge (system) pressure, differential pressure, and fluid temperature.

To sense temperature, a resistor is incorporated on the device's surface to function as a thermister. Establishment of the temperature-resistance characteristics of this thermister will enable temperature measurement. The design of the pressure sensing aspects of the MEMS device is more complex. First, the mechanical aspects of the sensor diaphragms will be considered. Once the diaphragm size and thickness has been established, consideration will be given to the optimum placement of the piezoresistive strain elements. The physical design of a MEMS device is closely coupled with the design of the fabrication procedure. The design of the physical components of the sensor will be presented first, followed by the fabrication sequence.

4.1   Thermister Design

The primary consideration in the design of a thermister is self-heating error, which appears directly as uncertainty in the temperature measurement. Therefore, design consideration needs to be given to both the power dissipated by the device and heat transfer. Measurement of thermister resistance in the PQT application uses a Wheatstone bridge (Fig. 18.19), with the thermister functioning as a single resistance element. The amount of energy dissipated by the thermister is a function of its resistance and the bridge excitation voltage.

In Fig. 18.19, if R_4 is the thermister, and the bridge is initially balanced, the power dissipated is a function of the thermister resistance and the excitation voltage. A bridge excitation voltage of 5 V is used for the MEMS thermister. If the thermister has a nominal resistance of $10 \, k\Omega$, the power is 0.625 mW. As a comparison, thermisters used to measure blood perfusion dissipate 4 mW of power, and can self-heat to $3°$ or $4°$ greater than ambient. This type of device is typically spherical, and not in direct contact with the fluid [3]. In this application, the thermister is fabricated on the surface of the device in a serpentine fashion, has a large surface area in direct contact with the fluid, and dissipates 85 % less power. Therefore, self-heating errors can be assumed negligible compared to the expected change in temperature of the hydraulic fluid.

Fig. 18.19 Standard wheatstone bridge
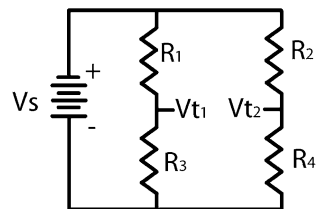
To simplify device fabrication, the thermister is built from doped polysilicon, the same piezoresistive strain material used for pressure sensing. Polysilicon has a positive temperature coefficient of resistance (TCR) of 0.08 %/°C (at doping levels of 10^{20} cm^3) [34]. Although other materials with order of magnitude higher TCR exist, they are more complicated to use in device fabrication.

4.2 Pressure Sensor Design

The functional principle for pressure sensing in the MEMS device can be seen in Fig. 18.20. A diaphragm is subject to a pressure, which causes a small deflection and strain on the surface. This strain can be converted to an electrical signal through the use of piezoresistors, or resistors whose resistance changes with applied strain. The change in resistance can then be converted to an output signal using a Wheatstone bridge. Through calibration, the bridge output signal is then related to the applied pressure.

The differential pressure is typically much smaller than the system pressure. The differential pressure sensor must therefore be much more sensitive than the system pressure sensor. To achieve the desired sensitivity with reasonable diaphragm dimensions, the diaphragm must have a thickness on the order of microns. Thickness control of silicon diaphragms of this size is difficult; therefore, an alternative material is needed. Because cost is of importance, silicon dioxide has been selected for use as the diaphragm material. The required sensitivity of the system pressure sensor is much lower, as it is subject to a much higher pressure difference ($\sim$ 4.0 MPa). This requires a thicker diaphragm, which can easily be fabricated from the base silicon material. Adjusting the thickness and size of the diaphragms would allow for design pressures other than those selected for this demonstration.
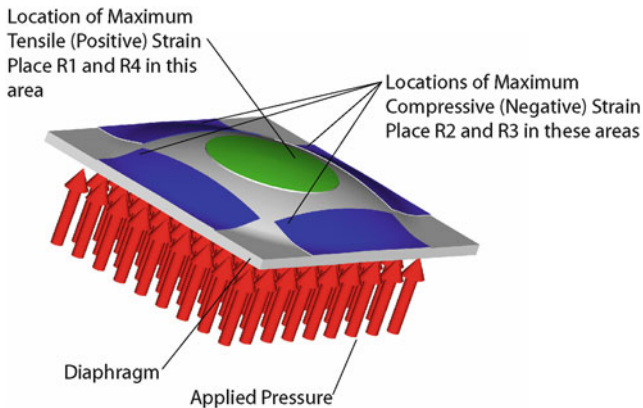
Fig. 18.20 Piezoresistive pressure sensor principle

The use of potassium hydroxide (KOH) to etch the bulk of the silicon required the diaphragms to be made square. Although circular diaphragms have better stress properties than square diaphragms, the stress advantages do not outweigh the increased fabrication costs to etch a circular diaphragm.

The pressure sensing theory of operation suggests a logical design path. First, the mechanical properties of the diaphragm need to be selected to maximize the strain at the surface, while providing a sufficient factor of safety against breakage. After the diaphragms have been properly sized, the location and size of the piezoresistors needs to be selected to maximize the electrical output of the device. Once the size and location of the piezoresistors are established, the physical location on the die of each functional component of the device can be selected. Finally, the layout of the bonding pads and wire traces can be done.

Sizing of the diaphragms was done by modeling them as plates of fixed periphery. Since the governing differential equation for a plate of this type has no closed form solution, the Ritz energy method was used to approximate the solution. The displacement of the differential sensor diaphragm (made of $2\,\mu m\ SiO_2$) can be seen in Fig. 18.21.

Comparing with solution given by a full finite element analysis, it is found that retaining only one term of the series expansion is adequate. The Ritz energy approximation for the plate of fixed periphery allowed the strain at the surface of the plate to be known at any location on the surface. The strain at the top surface of the differential pressure diaphragm can be seen in Fig. 18.22.
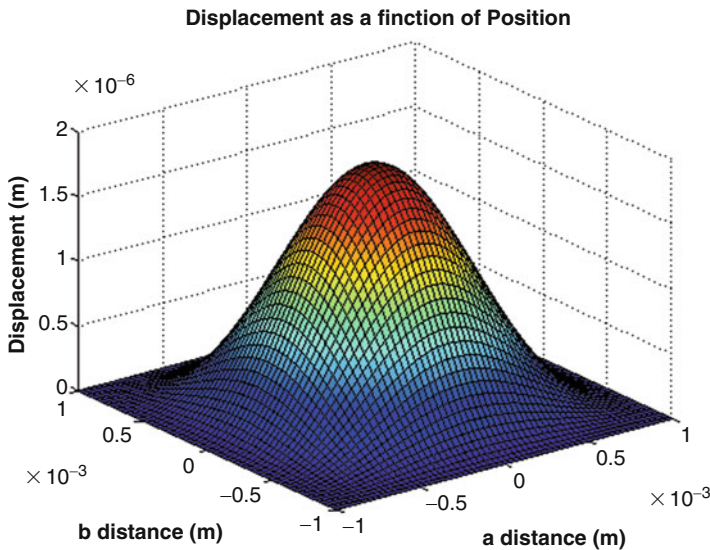
Fig. 18.21 Displacement of differential pressure sensor diagram given by Ritz energy method approximation to a plate of fixed boundary conditions subject to maximum design pressure

Strain in X Direction as a finction of Position
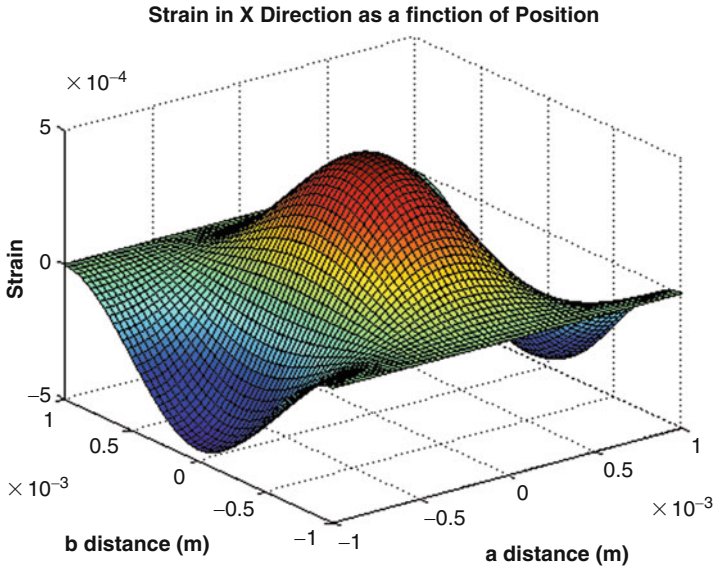
Fig. 18.22 Strain in a single direction as a function of position on the top surface

Fig. 18.23 Serpentine resistor layout
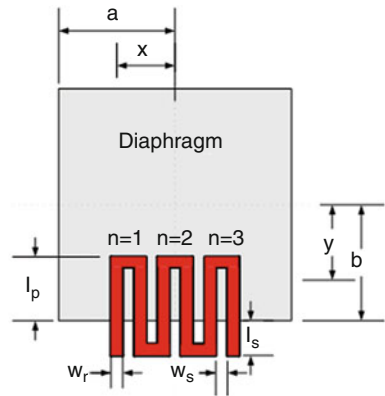
Since the fractional change in resistance of the piezoresistor is proportional to the applied strain, maximizing the applied strain also maximizes the change in resistance.

The piezoresistors were designed in a serpentine fashion and allowed to vary in both size and location on the surface of diaphragm. An illustration of such a resistor layout can be seen in Fig. 18.23.

To maximize the output of the sensor, each of the variables of the piezoresistors shown in Fig. 18.23 is systematically varied. The change in resistance is estimated by area averaging. To reduce the number of variables in the solution domain, the
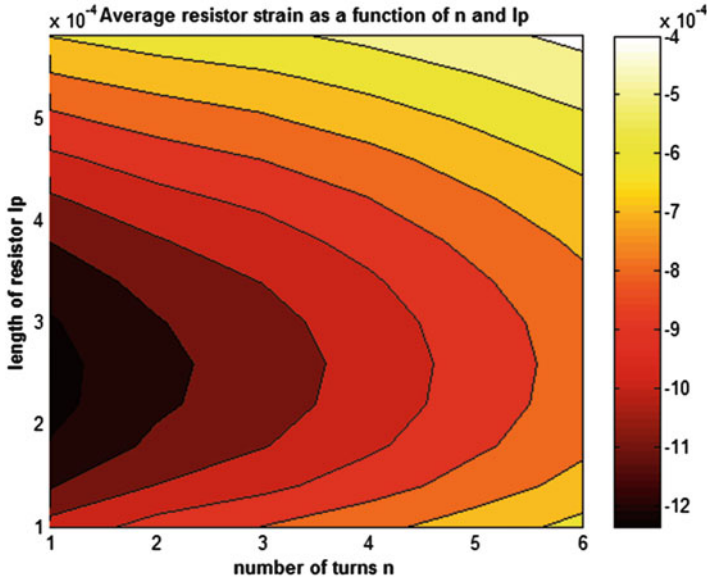
Fig. 18.24 Bridge output as a function of resistor geometry

length the resistor extended off of the diaphragm, l_s, is held constant at $100\,\mu$m. Similarly, the width of the spacing, w_s, and the width of the resistor, w_r, are held constant at $50\,\mu$m, a size that could be easily and cheaply fabricated. The output of a Wheatstone bridge of piezoresistors arranged in a full bridge pattern on the surface of the system pressure diaphragm is shown in Fig. 18.24. A single turn resistor which gives the maximum resistor output results in substantially degraded signal-to-noise ratios [35]. Therefore, a longer resistor with 2 turns and length, $l_p = 250\,\mu$m is chosen. The pressure sensor model underpredicts the output voltage. This is because the model considers only strain information on the diaphragm, and assumes the strain is zero at all locations off the diaphragm. The actual physical system does have strain off of the plate, resulting in a higher output signal.

A similar design procedure was followed for the design of the system pressure sensor. For this sensor, the diaphragm thickness is $90\,\mu$m. This requires the diaphragm size to be 1.6 mm. Because the diaphragm size is larger, 3 resistor turns are used. After the piezoresistors are placed on the diaphragms, the die layout is determined. For simplicity, the pressure sensors are placed in the vertical center of the die, equidistant from the horizontal centerline. This location maximizes the sealing surface around the system pressure sensor. Large 1.1×1.1 mm contact pads are chosen to allow connection to the sensor without wire bonding equipment. The thermister is placed off center between two contact pads to simplify wire trace layout. Large $80\,\mu$m wide wire traces are used for the wiring of sensor so they can be seen with the naked eye. The layout of the device is shown in Fig. 18.25.
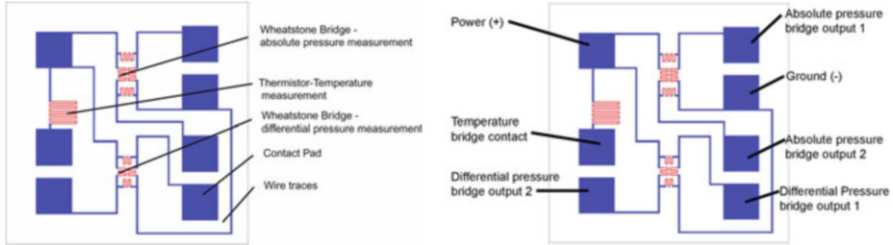
Fig. 18.25 MEMS sensor component and contact pad layout on die

4.3 Sensor Fabrication

Fabrication of the MEMS device uses standard deposition, lithographic and etching techniques to reduce cost. A basic overview of the MEMS device fabrication sequence is given in Fig. 18.26.

The fabrication process uses clean $525\,\mu$m thick [1 0 0] oriented double polished wafers. The resistivity or doping of the base silicon does not affect the device performance as all electrical aspects of the device are created with surface micromachining on an insulating oxide layer.

Fabrication starts by wet oxidizing the base wafers to form a $2\,\mu$m thick layer of SiO2 on both sides. Wet oxidation is used because it is a faster process, and also because the resulting films have much less residual stress than films produced by a dry oxidation process. However, dry oxidation processes result in higher film qualities, and the factor of safety in the differential diaphragm design was selected based on the use of the wet process [36, pp. 242–245, 345–349],[37, pp. 388–441].

Following the oxidation of the wafers, a $5000\,\overset{\circ}{A}$ thick layer of in-situ doped polysilicon was deposited using a standard low pressure chemical vapor deposition process (LPCVD). The in-situ process was used in lieu of a separate doping process to reduce cost and to simplify fabrication. Moreover, in-situ doping using the LPCVD process can be tailored to deposit films with no residual stresses [38, pp. 20–22]. A recrystallization post deposition annealing step was done to the polysilicon to reduce the sheet resistance. This process was performed in an inert N_2 environment to reduce the oxidation of the film. Following annealing, the polysilicon was masked using standard lithographic processes, and the bulk of the material etched away using reactive ion etching (RIE). The etched material included that of the backside, where doped polysilicon was not needed. After etching, the wafer appeared as in step c of the fabrication sequence shown in Fig. 18.26.

The doped polysilicon piezoresistors were connected using gold contact traces. Past researchers have noted adhesion difficulty in the application of gold directly to silicon dioxide. Therefore, a standard adhesion promoting $2000\,\overset{\circ}{A}$ thick chromium layer was deposited using electron beam evaporation prior to evaporation of a
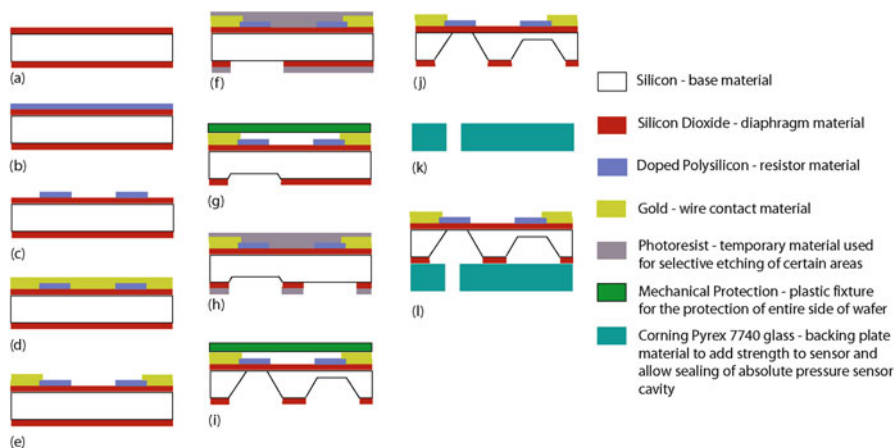
Fig. 18.26 Fabrication sequence. (**a**) Thermal oxidation; (**b**) LPCVD doped polysilicon; (**c**) Pattern polysilicon; (**d**) Evaporate gold; (**e**) Pattern gold; (**f**) Oxide Etch—differential pressure diaphragm; (**g**) KOH Etch silicon—differential pressure cavity; (**h**) Oxide Etch—absolute pressure diaphragm; (**i**) KOH Etch silicon—both pressure sensor cavities; (**j**) Finished Si component; (**k**) Cut and drill Pyrex 7740 glass; (**l**) Anodic bonding of Si to glass

3500 Å thick gold layer. Following deposition, a lithographic step masked the wire traces and bonding pads, and the bulk of the gold, and chromium was wet etched from the wafer surface to form both the wire traces and the contact pads. Following gold deposition and patterning, the wafer now appears as in step e of Fig. 18.26, with all topside fabrication is completed.

For ease of prototype fabrication, a timed etch stop technique was used to control the thickness of each of the sensor diaphragms. Therefore, in order to size the thickness of the system pressure diaphragm correctly, and to etch through the wafer for the correct sizing of the differential pressure diaphragm, the etching of the differential cavity was started first. Prior to etching the base silicon material, a hole was created in the silicon dioxide on the backside to expose it. This was done using buffered oxide etch (BOE) solution and standard lithographic means. Upon creation of a window in the oxide, potassium hydroxide (KOH) was used for the bulk etching of the Si. Since the silicon dioxide is being used as the masking layer for this etch, care was given to the KOH bath temperature and concentration to ensure sufficient selectivity [39–41].

Care was also needed to protect the previously fabricated topside features during the KOH etching of the backside. Through experimentation, a combination of mechanical and chemical means is determined to provide sufficient protection. Mechanical protection is provided with a plastic fixture with an o-ring which seals the topside features in a liquid tight chamber. Since a pinhole or diaphragm breakage would still result in KOH solution reaching the sensitive topside components, a chemical barrier was also applied [42]. After bulk micromachining to form the

pressure sensor diaphragms, the finished Si sensor component was to be anondically bonded to Pyrex 7740 glass. This process has been successfully applied in the commercial production of many sensors [43]. However, for the purpose of prototype testing, an alternative, faster method was used. In this method, the finished Si sensor is bonded to a blank Si backing plate with allowance for the differential pressure signal using Loctite epoxy based adhesive. The epoxy is adequate for sealing against pressures that the sensor would see.

Several challenges were encountered during the fabrication of the MEMS PQT device. These challenges were substantial enough to slightly alter the planned fabrication sequence and deserve mention.

The purpose behind fabricating prototype sensors was to demonstrate the feasibility of the integrated PQT concept. Therefore, fabrication issues such as repeatability among devices on a wafer were of low concern. This allowed cheaper masks fabricated from high resolution transparencies to be used. The fabrication of these masks limits the minimum usable feature size to $50 \, \mu m$, and the feature size variation from device to device is not negligible. The advantage to using such masks is a very fast turnaround time, and $\sim 66 \%$ saving compared to CNC produced masks. All masks used in the fabrication of the prototype MEMS sensors were of the printed transparency type.

Difficulty was found in the fabrication of the differential pressure sensor diaphragm. Silicon dioxide has considerable compressive residual stress after deposition [36, 37]. During prototype fabrication, it was determined this residual stress was high enough to cause breakage of the thin diaphragm once all of the silicon backing material had been etched away. Several attempts were made to anneal and relax this residual stress, however they were unsuccessful. Success in fabricating this diaphragm was found through a combination of deep reactive ion etching (DRIE) and silicon support. The technique was to follow the standard etching steps until there was approximately $40 \, \mu m$ of material remaining. Then, DRIE was used to etch away approximately $25 \, \mu m$ of material, leaving $15 \, \mu m$ of material to support the SiO2 diaphragms. The use of the DRIE to finish etch allowed radii to be formed on the corners in the pressure sensor cavities, which eliminated the sharp edge stress concentrations that would ordinarily have been present if the KOH etch was finished. The reduction in the stress concentration along with the support gained by the remaining silicon material was sufficient to produce operational and sensitive differential pressure sensors.

During the initial phases of fabrication, connection difficulties between the Cr/Au wires and the piezoresistors were encountered. This difficulty resulted in very low batch yields, and operational sensors that produced very erratic output signals. The difficulty was traced to poor cleaning of the polysilicon layer before the e-beam evaporation of the Cr/Au layers. Since the polysilicon was both annealed, and patterned using standard lithographic techniques, there was a thin oxide and residual photoresist present. An intermediate RIE etch prior to the evaporation of the Cr/Au layers provided the necessary cleaning.

5   Sensor Testing and Evaluation

5.1   Thermister Evaluation

The performance of the thermister was characterized through comparison with a K calibration thermocouple using an Analog Devices AD595 thermocouple signal conditioning chip. For the characterization tests, the MEMS PQT device and the thermocouple were placed in an oil bath. The bath was heated and allowed to cool, while both the resistance of the MEMS PQT device and the temperature of the bath were monitored.

An example calibration curve for several thermal cycles on a single sensor is shown as Fig. 18.27. Note the temperature–resistance characteristic is linear over the small temperature. Therefore, a Steinhart-Hart polynomial log fit is unnecessary. The temperature coefficient of resistance can be obtained from the trend line slope indicated in Fig. 18.27. From this slope, a temperature coefficient of resistance for the doped polysilicon used in the MEMS PQT device is calculated to be 0.08 %/°C, which matches the value given by Madou [34].

When the calibration curve shown in Fig. 18.27 is used, the performance of the thermister can be evaluated. The result is shown graphically in Fig. 18.28. The thermister exhibits a linearity of 2.32 %, a repeatability of 0.6 %, and an overall accuracy of 3.6 %. The percentage accuracy of the device translates to a 1.5 °C uncertainty in temperature measurement with the device.

5.2   Differential Pressure Evaluation

Testing of the differential pressure sensor required a holding fixture to be fabricated. There are two pressure taps in the fixture allowing for a differential signal to be applied to the MEMS device.
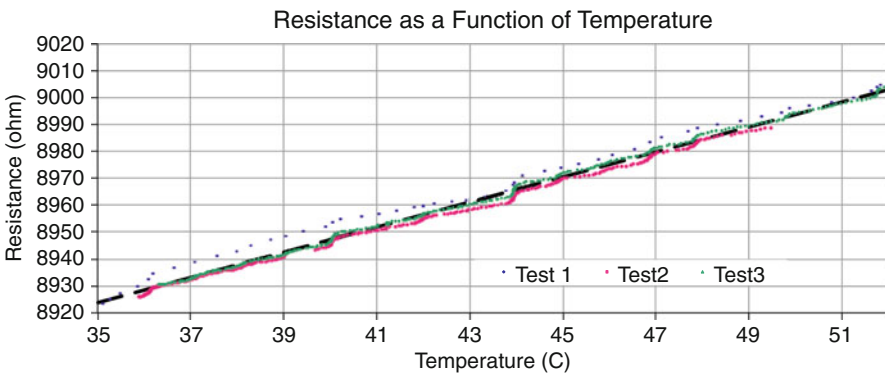
Fig. 18.27   Thermister temperature resistance curve
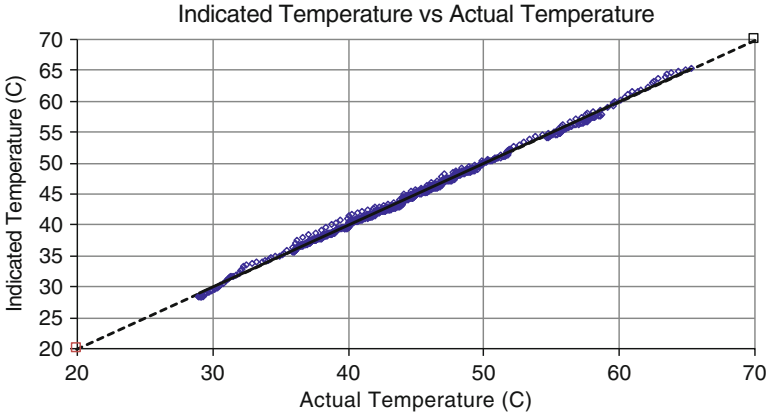
Fig. 18.28 Thermister performance curve
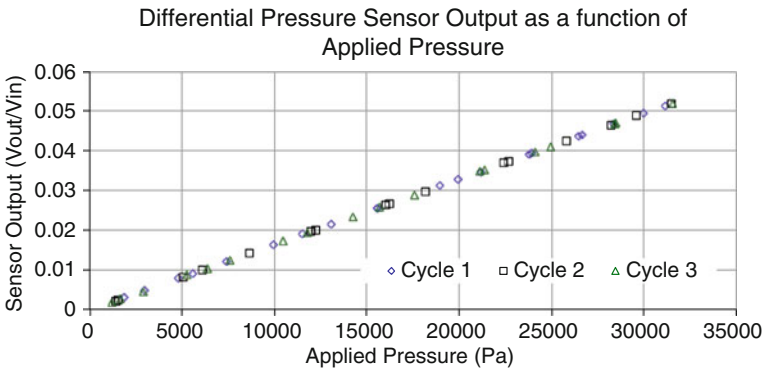
Fig. 18.29 Pressure-voltage characteristic of the differential pressure sensor

A differential pressure signal was obtained by subjecting one side of the MEMS differential pressure sensor to atmospheric pressure, while the other side was subjected to the pressure generated by a variable head of hydraulic oil. The differential pressure measurement was measured with a high accuracy Sensotec LVT differential pressure transducer. The results of the offset corrected performance test can be seen in Fig. 18.29.

The linearity of the device was measured to be 0.38 %, the repeatability 0.13 % and the accuracy 3.58 %. Further increases in the sensitivity through a reduction in the amount of silicon support under the differential diaphragm resulted in undesirable nonlinear output characteristics.

Comparing the experimental results in Fig. 18.29 with the model prediction in Fig. 18.24, it is seen that the model underpredicted the device output. However, the results of the model were comparable to the sensor output, which validates the use of the model for optimizing the output of the sensor.

5.3 System Pressure Evaluation

Performance testing of the system pressure component of the prototype MEMS device was very similar to that of the differential pressure tests. For this test, the output signal of the MEMS device was compared to a commercially available Barksdale pressure transducer. To avoid breakage of the differential pressure diaphragms during the tests, a high pressure hydraulic signal was applied to both sides of the test fixture and MEMS device.

Figure 18.30 shows the performance of the system pressure measurement. System pressure measurement exhibited a linearity of 0.75 %, a repeatability of 0.33 %, and an accuracy of 4.2 % when the turndown was restricted to 10:1. The performance is comparable to commercially available pressure transducers. Slightly degraded performance was noted when measuring small pressure signals outside of the turndown range.

The results of the tests of the three sensors: thermister, differential pressure sensor, and system pressure sensor are summarized in Table 18.2.
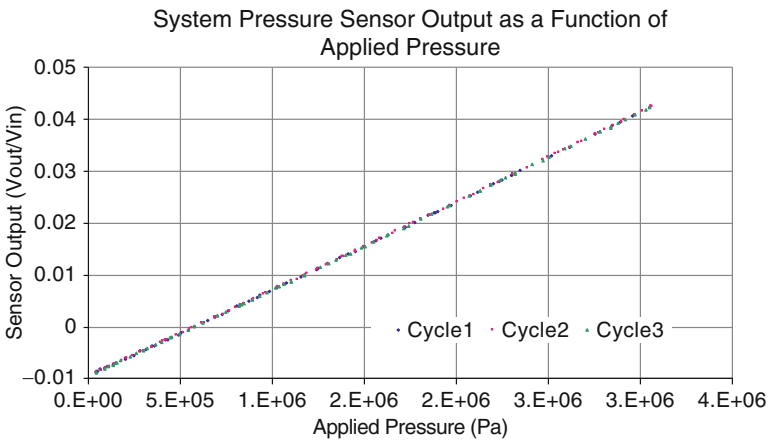
Fig. 18.30 Pressure voltage characteristic of system pressure sensor

Table 18.2 MEMS device performance summary

MEMS device	Thermister (%)	Diff. pressure (%)	System pressure (%)
Linearity	2.32	0.38	0.75
Accuracy	3.6	3.58	4.2
Repeatability	0.60	0.13	0.33

5.4 Sensor Integration into Elbow Geometry

Integration of the MEMS device into a typical fluid power component is necessary to prove the validity of the integrated PQT concept. The design is shown in Fig. 18.31. This design incorporates a standard 1/4 " (6.35 mm) flow bend with a curvature to diameter ratio of unity, and the MEMS device. The overall dimensions are 25 mm × 25 mm × 30 mm. The cavity housing the MEMS sensor in the example is 18 mm × 18 mm × 6 mm. Pressure taps are placed at 55° on the outer radius and at 25° on the inner radius for increased signal sensitivity.

Integrating the MEMS device into the fixture shown in Fig. 18.31 allows all sensing aspects of the device to be tested simultaneously. Since the temperature and system pressure sensor have already been evaluated in the previous sections, the remaining task is to evaluate flow sensing of the sensor. Application of the calibration procedure developed in Sect. 3 to the design shown in Fig. 18.31 results in the flow performance curves using the two different sets of calibration variables as shown in Fig. 18.32. Note that Fig. 18.32 contains experimental data collected at temperatures of 27.2 °C, 34.1 °C, 41.3 °C, and 43.5 °C. Linearity, accuracy, and repeatability using the 2 calibration methods are summarized in Table 18.3.

From Fig. 18.32, both calibration methods are able to collapse data from different temperatures onto a single curve. From Table 18.3, calibration method using π_1

Fig. 18.31 A machined elbow with pressure taps into which the PQT sensor is to be integrated

Table 18.3 MEMS device flow performance summary

Calibration variables	π_1, R_e (%)	Eu_{mod}, R_e (%)
Linearity	2.83	5.3
Accuracy	2.8	6.0
Repeatability	0.92	1.26
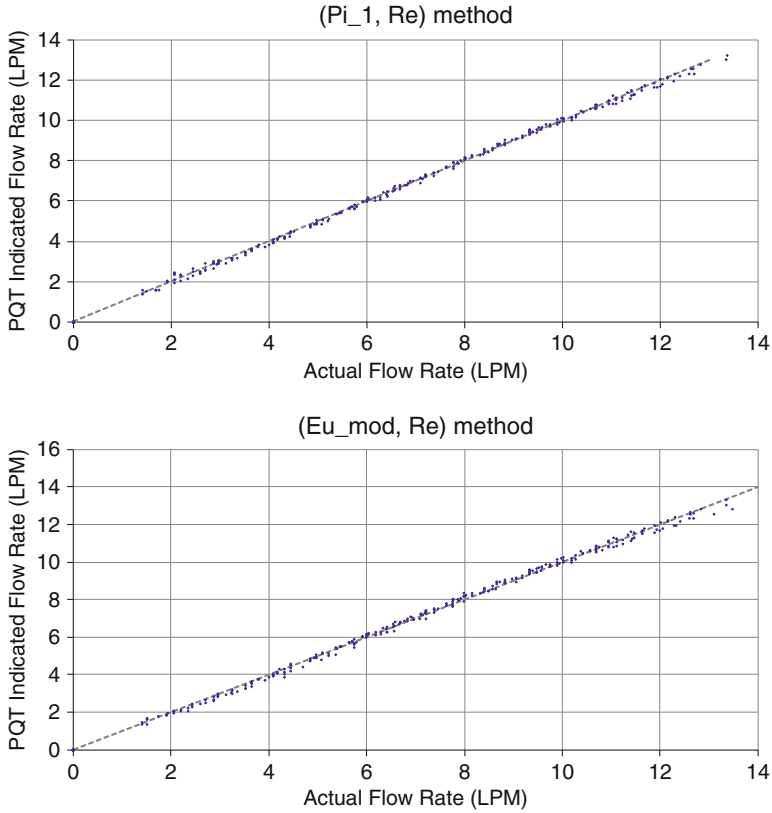
Fig. 18.32 PQT flow rate performance curves: *Top*: using π_1 and R_e as calibration variables; *Bottom*: using Eu_{mod} and R_e as calibration variables

and R_e as calibration variables shows better performance than the method that uses Eu_{mod} and R_e as calibration variables. This result is consistent with the result in Sect. 3.

6   Conclusions

The design of a low cost, multifunction MEMS sensor for use in fluid power systems was presented. The sensor measures system temperature, system pressure, and a differential pressure for calculating flow rate.

Temperature sensing with the device was accomplished by incorporating a thermister into the design. For fabrication simplicity, polysilicon was used as the thermister material. Although this material does not exhibit large resistance

temperature characteristic of other materials, the fabrication simplicity gained justified its use.

Pressure sensing with the device was accomplished by using piezoresistive strain elements arranged in a full bridge fashion. A simple fixed plate diaphragm model was used to optimize both the mechanical diaphragm properties and the placement of the piezoresistors on the diaphragm surface.

Flow sensing with the device was accomplished by sensing the difference in pressure at two locations in an existing hydraulic flow path of the application. This eliminates the need for additional elements in the flow path and is more energy efficient. Two calibration methods were developed based on dimensional analysis to determine the differential pressure-flow relations that compensate for viscosity and density variation due to temperature changes. The method that uses π_1 and R_e as calibration variables is more accurate than the method that uses Eu_{mod} and R_e.

Independent testing of each of the component on the device showed each has good functionality and performance characteristics. The uncertainty in the temperature measurement was determined to be 1.50 C, while the linearity of the system pressure and differential pressure measurements of the device were determined to be 0.13 and 0.75 %, respectively. The MEMS device was incorporated into a typical fluid power component proving the validity of the integrated PQT sensor. The linearity of the flow portion of the integrated PQT concept was determined to be 2.83 %. Further details of this work can be found in [44].

References

1. Bistue G, Elizalde JG, Garcia-Alonso I, Olaizola S, Castano E, Gracia FJ, Garcia-Alonso A (1997) A micromachined pressure sensor for biomedical applications. J Micromech Microeng 7:244–246
2. Boillat M, Van Der Schoot B, Droz B (2000) A high-precision piezoresistive flow sensor for microdosing applications. Sensors 17(9)
3. Chevroulet M (2000) A pressure sensor for a smart barometer. Sensors 17(5)
4. Eaton W, Smith J (1997) Micromachined pressure sensors: review and recent developments. Smart Mater Struct 6:530–539
5. Fung C, Harris R, Zhu T (1999) Multifunction polysilicon pressure sensors for process control. Sensors 16(10)
6. Muntjes JA, Hafner J, Gortz M, Mokwa W (2013) Studies on thinned flexible integrated capacitive pressure sensors in tactile sensor arrays for the use in robotics and prosthetics. In: Transducers Eurosensors XXVII: the 17th International Conference on Solid-state sensors, actuators and microsystems, pp 1460–1463
7. Fan F-R, Lin L, Zhu G, Wu W, Zhang R, Wang ZL (2012) Transparent triboelectric nanogenerators and self-powered pressure sensors based on micropatterned plastic films. Nano Lett 12(6):3109–3114. PMID: 22577731

8. Kim A, Powell CR, Ziaie B A generic packaging technique using fluidic isolation for low-drift implantable pressure sensors. In: Transducers - 2015 18th international conference on solid-state sensors, actuators and microsystems (Transducers), pp 476–479

9. Shahmohammadi M, Souri K, Makinwa KAA (2013) A resistor-based temperature sensor for MEMS frequency references. In: 2013 Proceedings of the ESSCIRC, pp 225–228

10. Souri K, Chae Y, Makinwa K (2012) A CMOS temperature sensor with a voltage-calibrated inaccuracy of 0.15c from 55 to 125 c. In: 2012 IEEE international solid-state circuits conference digest of technical papers (ISSCC) pp 208–210

11. Harada S, Honda W, Arie T, Akita S, Takei K (2014) Fully printed, highly sensitive multifunctional artificial electronic whisker arrays integrated with strain and temperature sensors. ACS Nano 8(4):3921–3927. PMID: 24580035

12. He RX, Lin P, Liu ZK, Zhu HW, Zhao XZ, Chan HLW, Yan F (2012) Solution-gated graphene field effect transistors integrated in microfluidic systems and used for flow velocity detection. Nano Lett 12(3):1404–1409. PMID: 22324366

13. Lotters JC, Groenesteijn J, van der Wouden EJ, Sparreboom W, Lammerink TSJ, Wiegerink RJ (2015) Fully integrated microfluidic measurement system for real-time determination of gas and liquid mixtures composition. In: Transducers - 2015 18th international conference on solid-state sensors, actuators and microsystems (Transducers), pp 1798–1801

14. Li P, Lei N, Sheadel DA, Xu J, Xue W (2012) Integration of nanosensors into a sealed microchannel in a hybrid lab-on-a-chip device. Sensors Actuators B Chem 166–167(Complete):870–877

15. Webster JG (2000) Mechanical variables measurement. CRC, Boca Raton

16. Addison H (1938) The use of pipe bends as flow meters. Engineering 227–229

17. Dean W (1927) Note on motion in a curved pipe. Philos Mag Lett 4(20):208–223

18. Dean W (1928) The streamline motion of fluid in a curved pipe. Philos Mag Lett 5(30):673–695

19. Dean W (1928) Fluid motion in a curved channel. Proc R Soc Lond A121:402–420

20. Einhellig R, Schmitt C, Fitzwater J (2002) Flow measurement opportunities using irrigation pipe elbows. In: Proceedings of the ASCE specialty conference - hydraulic measurements and experimental methods

21. Eustic J (1910) Flow of water in curved pipes. Proc R Soc Lond 84(568):107–118

22. Lansford W (1934) Use of an elbow in a pipe line as a means of measuring the flow of water. Bull Assoc State Eng Soc 9(2)

23. Lansford W (1936) The use of an elbow in a pipe line for determining the rate of flow in the pipe. Bull Univ Ill Eng Exp Stn Urbana Ill 289:5–33

24. Levin A (1914) A flow metering apparatus. Trans Am Soc Mech Eng 36:239–254

25. Nishiyama S et al. (1999) Hydraulic design of bypass flow meter using the bend pipe. In: Proceedings of the international conference on agricultural engineering, pp 247–254

26. Taylor D, McPherson M (1954) Elbow meter performance. J Am Water Works Assoc 46:1087–1095

27. Taylor G (1929) The criterion for turbulence in curved pipes. Proc R Soc Lond 124(794):243–249

28. Yarnell D, Nagler F Flow of water around bends in pipes. Trans ASCE 100:1018–1043

29. Matsuoka Y, Yamamoto Y, Yamada K, Shimada S, Tanabe M, Yasukawa A, Hidetoshi M (1994) Characteristic analysis of a pressure sensor using the silicon piesoresistance effect for high pressure measurements. J Micromech Microeng 5:25–31

30. Fox RW, McDonald AT (1998) Introduction to fluid mechanics, 5th edn. Wiley, New York

31. Munson BR, Young DF, Okiishi TH (2002) Fundamentals of fluid mechanics, 4th edn. Wiley, New York

32. Shaw R (1960) The influence of hole dimensions on pressure measurements. J Fluid Mech 7(4):550–564

33. Buckingham E (1914) On physically similar systems: illustrations of the use of dimensional equations. Phys Rev 4:345–376

34. Madou M (2002) Fundamentals of microfabrication. CRC/LLC, Boca Raton

35. Bae B, Flachsbart BR, Park KK, Shannon MA (2004) Design optimization of a piezoresistive pressure sensor considering the output signal-to-noise ratio. J Micromech Microeng 14:1597–1607
36. Alexandrova S, Szekeres A, Christova K (1988) Stress in silicon dioxide films. Philos Mag Lett 58(1):33–36
37. Yu C, Flinn P, Bravman J (1997) In-situ measurement of viscous flow of thermal silicon dioxide thin films at high temperature. In: Proceedings of the 1997 material research society symposium, no. 446, pp 261–266
38. Oei D, Mccarthy S (1992) The effect of temperature and pressure on residual stress in lpcvd polysilicon films. In: Proceedings of the 1992 material research society symposium, no. 276, pp 85–90
39. Kovacs G, Maluf N, Peterson K (1998) Bulk micromachining of silicon. Proc IEEE 86(6):1536–1551
40. Merlos A, Santander J, Alvarez MDMD, Campabadal F (2000) Optimized technology for the fabrication of piezoresistive pressure sensors. J Micromech Microeng 10:505–510
41. Seidel H, Csepregi L, Heuberge A, Baumgartel HH (1990) Anisotropic etching of crystalline silicon in alkaline solutions. Parts 1 and 2. J Electrochem Soc 137(11):3612–3635
42. Ruben K, Flaim T, Li C (2004) Polymeric protective coatings for MEMS wet-etch process. In: Proceedings of SPIE: micromachining and microfabrication process technology, no. 5342
43. Mirza A, Ayon A (1998) Silicon wafer bonding: key to MEMS high volume manufacturing. Sensors 15:24–33
44. Groepper C (2006) Integrated pressure, flow and temperature sensing for fluid power systems. Master's thesis, Department of Mechanical Engineering, University of Minnesota

Chapter 19
Vibrating Nanoneedle for Single Cell Wall Cutting

Md. Habibur Rahman, Abdul Hafiz Mat Sulaiman, Mohd Ridzuan Ahmad, Masahiro Nakajima, and Toshio Fukuda

Abstract In this chapter, we proposed a piezoelectric-actuated rigid nanoneedle for single cell wall (SCW) cutting. A fabricated Tungsten (W) nanoneedle was assembled with a commercial piezoelectric actuator laterally and perpendicularly. The body of the nanoneedle is cylindrical while the shape of the tip is conical. Calibration of the assembled device was conducted using both experimental and numerical approaches. Experimentally, we have obtained a 4.5 µm displacement of the nanoneedle from an input potential of 100 V towards the *Y*-axis. Results also showed that, at lateral configuration, nanoneedle tip is more stable than at perpendicular configuration. For the perpendicular configuration, we observed a fractional displacement of 1.2 pm, but in the lateral configuration we did not experience any fractional displacement. As a result, for a precise cell wall cutting we choose the lateral configuration of the nanoneedle tip. W303 yeast cell model was used as a sample for SCW cutting. Yeast cell wall is more likely to behave as a Mooney–Rivlin hyperelastic material. Two different frequencies of vibrating nanoneedle were used for cell wall cutting operation: 1 and 10 Hz. For a constant penetration depth of 1.2 µm, penetration velocity was 7 µm/s and 24 µm/s, respectively. Investigated results showed that the faster nanoneedle causes less damage to the cell wall during cutting operation. Additionally, further finite element

M.H. Rahman
Department of Computer Science and Engineering, University of Asia Pacific, Green Road, Dhaka-1215, Bangladesh

A.H.M. Sulaiman • M.R. Ahmad (✉)
Department of Control and Mechatronics Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
e-mail: ridzuan@fke.utm.my

M. Nakajima
Center for Micro-nano Mechatronics, Nagoya University, Nagoya, Aichi, Japan

T. Fukuda
Institute for Advanced Research, Nagoya University, Nagoya, Aichi, Japan

Faculty of Science and Engineering, Meijo University, Nagoya, Aichi, Japan

Intelligent Robotics Institute, School of Mechatronic Engineering, Beijing Institute of Technology, Beijing, China

(FE) analysis was performed to observe the effect of tip edge angle for cell wall cutting. Three different edge angles 10, 15, and 20° were used to perform cell wall cutting. SCW penetration force increases with the increasing of tip edge angle and the diameters of the nanoneedle.

Keywords Vibrating nanoneedle • Single cell cutting • Piezoelectric actuator • Finite element analysis • Yeast cell model

1 Introduction

The fundamental structure of a living organism is cell. Millions of cells are combined together to develop a total structure of tissue. Therefore single cell analysis is very important in tissue engineering. Single cell analysis elucidates complex cellular functions such as cell's mechanical, electrical, and chemical properties.

Conventional medical science researches are based on a population cell analysis that produces a conclusion from an average data. However, the average data is not able to illustrate the basic physiological properties of cell such as cell membrane stiffness, cell wall thickness at different cell growth, and cell proliferations [1]. For instance, abnormal cell growth caused cancer or tumor [2, 3] by which intracellular and extracellular mechanical properties change significantly [4, 5]. From the biochemical experiments, it might be possible to identify that the cell growth is abnormal, but to identify the exact mechanical properties of the cell wall for a healthy and unhealthy cell, it is necessary to analyze cell's mechanical property individually.

Traditionally, diamond and glass knives were used for ultrathin cryosectioning of cells [6, 7]. But due to the sturdy edge of diamond knife and high edge angle (40–60°), it generates a very high compression stress on the upper surface of cells. This may lead the sample's inner surface to be cracked. Moreover, cryosectioning has been performed in a very low temperature or in a frozen-hydrated sample. On the other hand, cell mechanics change with environment temperature [8]. Therefore, ambiguities arise due to study during or after cell cutting operation.

Conventional diamond nanoknife was replaced with Multiwalled Carbon Nanotube (MWCNT)-based nanoknife for cell cutting [9]. A Carbon Nanotube (CNT) was welded across two tungsten needles to fabricate the nanoknife. Due to very small diameters of the CNT (100 NM), the compression force on the cutting sample surface has been reduced. Hence, the sample surface bent instead of cracking during cell cutting. It is seen clearly from the reported scanning electron microscopic (SEM) image that nanoknife can leave a mark on the Au-coated epon resin surface, which means it is able to generate very thin slices of the sample cell. However, electron-beam-induced deposition (EBID) bonding strength between CNT and Tungsten needles is 0.81 ± 0.8 μN. As a result, continuous cell cutting operation is not possible by CNT nanoknife particularly if sample cell is hard, like

plant cells, bones, thick walled spores, etc. [10, 11]. Recently, our colleagues have fabricated a novel nanoknife by FIB etching of a commercial AFM [12] cantilever. The nanoknife edge angle was 5° and the knife was supported by a buffering beam. Operation of cell cutting was performed inside environmental scanning electron microscopy (ESEM). *Saccharomyces Cerevisiae* yeast cell was used as a sample of cell cutting. Cutting force was measured as 60 μN for a yeast cell. However, this work was limited to single cell slice generation only. The reported data is not adequate enough to explain the stiffness of the single cell wall (SCW). The mechanical properties of the cell wall are partially extracted and yet under the area of near darkness [13]. Unlike the conventional methods of single cell slice preparation, our aim is to cut the SCW specifically. Therefore, in this chapter, we presented a piezoelectric-actuated vibrating rigid nanoneedle for SCW cutting. This approach of SCW cutting can be useful for early disease diagnosis based on cell's mechanical properties and also for drug delivery to the individual cell.

2   Proposed Idea

In this work, a vibrating rigid nanoneedle has been developed for SCW cutting. The vibration of the nanoneedle was controlled by a piezoelectric actuator's signal input. Figure 19.1 showed the concept of the SCW cutting operation. The inverse piezoelectric effect was used to control the vibration of the nanoneedle. According to the inverse piezoelectric effect, the external electric energy will generate internal mechanical energy or displacement [14]. Displacement can be controlled by applying voltage or current [15]. This internal mechanical energy will cause the actuator to displace from its original position. In our model, the nanoneedle is located at the tip of a commercially available AFM cantilever (Olympus Corp.). The body of
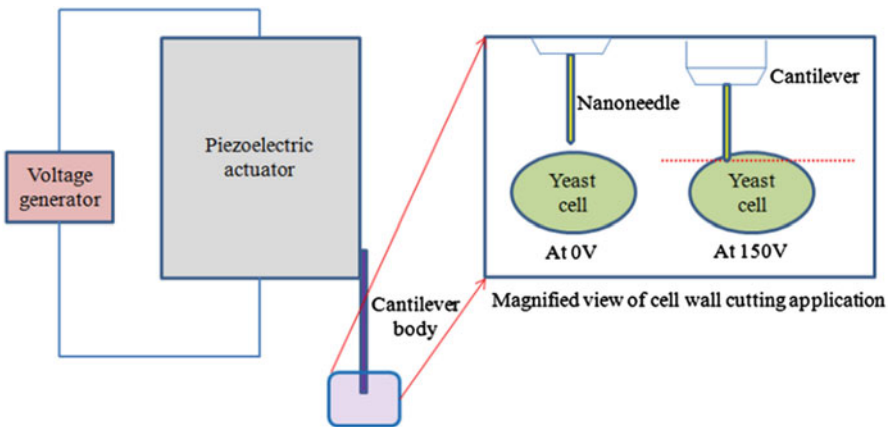
Fig. 19.1 Single cell wall cutting with piezoelectric-actuated rigid nanoneedle

the AFM cantilever is attached to the piezoelectric actuator (AE0203D04F, NEC Corp.). SCW penetration force was related with Hertz–Sneddon model of force vs. cell wall penetration depth [16].

2.1 Finite Element Model of the Fabricated Rigid Nanoneedle

Finite element (FE) method is a widely accepted tool for modelling and validating complex biotechnological analyses [17–19]. Commercially available powerful finite element software ABAQUS 6.10 CAE/CEL was used to model and analyze the cell wall cutting operation. The physical units of the numerical model were scaled to millimeters to fit in the modelling tools. Final results were obtained by multiplying by 10^{-03} to tally with the actual physical dimension. Nanoneedle is the end-effector to perform SCW cutting operation. The body shape of the nanoneedle is cylindrical while the tip shape of the nanoneedle is conical. The conical-shape tip of the nanoneedle increases its sharpness and this criteria is very important to obtain precise cell cutting with minimal damage to the cell wall [1].

Previously, Ahmad et al. had fabricated tungsten (W) deposited nanoneedle for SCW penetration and it was reported that the nanoneedle is able to penetrate the cell wall successfully [8]. Nanoneedle was fabricated in two different configurations: lateral and perpendicular. Figure 19.2a shows the lateral configuration of the fabricated nanoneedle and Fig. 19.2b shows the perpendicular configuration of the nanoneedle.

In this work, we have integrated the fabricated nanoneedle with lead zirconate titanate (PZT) actuator to perform cell wall cutting operation. FE method was used to analyze the data. The dimensions of the modeled nanoneedle are shown in Fig. 19.3. Figure 19.3a shows the schematic of the nanoneedle where its tip length is 350 nm and body length is 10 μm. Figure 19.3b illustrates a side view of the nanoneedle tip, where angle of the tip is 15°. The diameter of the nanoneedle is 170 nm as shown in Fig. 19.3c. Figure 19.3d shows the FE model of the whole
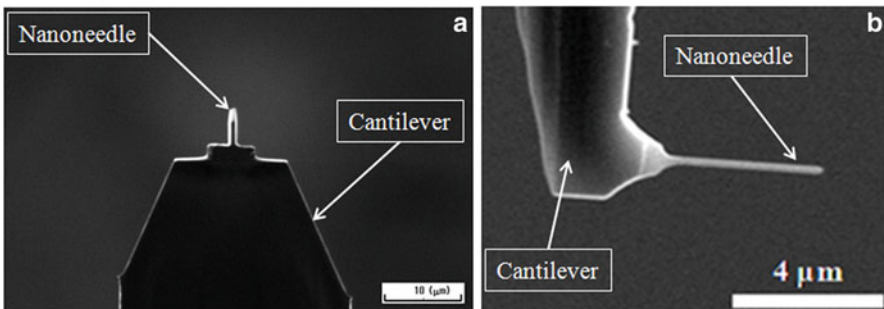
Fig. 19.2 (**a**). Lateral configuration of the nanoneedle. (**b**). Perpendicular configuration of the nanoneedle
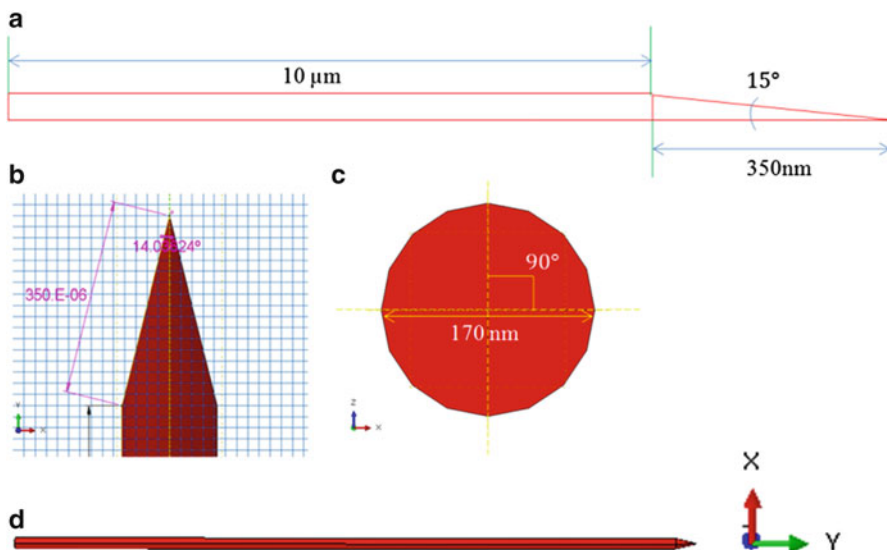
Fig. 19.3 (**a**). Schematic diagram of the nanoneedle. (**b**) Side view of the nanoneedle tip. (**c**). Cylindrical top view of the nanoneedle and its diameter. (**d**). Finite element structure of the nanoneedle

nanoneedle structure. The material of the nanoneedle is tungsten (W), which is a strong metal with high stiffness. The Young's modulus of tungsten is 411 GPa and its Poisson's ratio is 0.28 with a molecular mass density of 1.9 g/m^3 [20]. On the other hand, the stiffness of the cell wall is dependent on its basic properties such as Young's modulus, Poisson's ratio, and cell wall thickness. Yeast cell wall thickness is about 300 nm, Poisson's ratio 0.5, and Young's modulus 3.21 MPa, which is very low compared to the fabricated nanoneedle. In earlier work [8], it is reported that the fabricated nanoneedle is able to penetrate yeast cell wall without any buckle or any damage to the needle. In this current work, the same material was used to develop the nanoneedle, by which, we could claim that, the nanoneedle is rigid in terms of yeast cell wall cutting.

2.2   Piezoelectric Actuator

In the early history of piezoelectric properties, it was used to generate ultrasonic sound [21] in order to detect any object. In modern science, piezoelectric properties have been utilized in many areas, especially in micro- and nano-sensors. Its main advantages are the micro level displacement and low power consumption of the piezo material. Lead zirconate titanate (PZT) film was used by Kanda et al. [22] to measure the sensitivity of a sensor probe. Various types of ultrasonic motors have been also developed using the PZT film sensor probe. Various types of ultrasonic motors have been also developed using the PZT film.

Fig. 19.4 Concept of piezoelectric and inverse piezoelectric effect. In inverse piezoelectric effect electrical energy converts to mechanical energy in terms of displacement
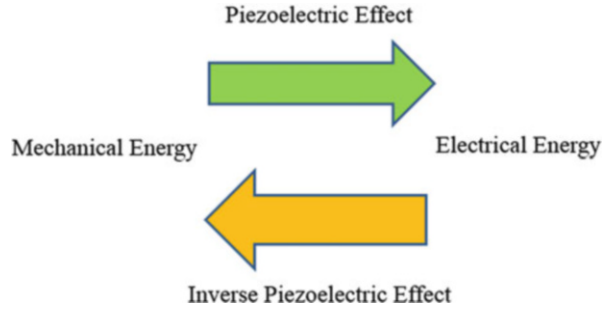
Table 19.1 Finite element model parameters for PZT actuator

Category	Parameters	Units
Elastic properties	$E = \begin{pmatrix} 6061 & 0.5 & 5300 \\ 8831 & 0.2 & 4350 \\ 9061 & 0.3 & 3300 \end{pmatrix} \times 10^{10}$	Pa
Permittivity	$\in = \begin{pmatrix} 1.5 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 1.3 \end{pmatrix} \times 10^{-08}$	F/m
Piezoelectric parameters	$d_{i,j,k} = \begin{pmatrix} 0 & 0 & 0 \\ -6.5 & 2.33 & 6.5 \\ 0 & 0 & 0 \end{pmatrix} \times 10^{-8}$	Cb/m

Piezoelectric ceramic is divided into two categories, i.e., Piezoelectric Effect and the Inverse Piezoelectric Effect. For developing a sensor and ultrasonic sound wave, researchers are using piezoelectric effect. On the other hand, the inverse piezo-electric effect has been used to develop actuators and vibrators [14]. Figure 19.4 shows the concept of piezoelectric effect. In this work, we have applied the inverse piezoelectric effect to the vibrating nanoneedle. The dimension of the piezoelectric actuator is 4 mm × 3 mm. Table 19.1 described the PZT properties of the actuator. These properties contain elastic, dielectric, and piezoelectric parameters for the actuator. The matrices were developed based on the elastic properties in FE model. It is clear that piezoelectric inputs were defined in Y direction, which limits the actuation in Y direction only. As our concern is in the vibration towards Y-axis, we kept other direction as static.

2.3 Assembling of the Nanoneedle and the PZT Actuator

We have assembled the nanoneedle with PZT actuator by keeping consistency with the nanoneedle configurations. For lateral configuration of the nanoneedle, cantilever was attached with the PZT actuator laterally, i.e., angle between PZT side and cantilever frame is 180°. Figure 19.5a shows the lateral configuration of the assembly.
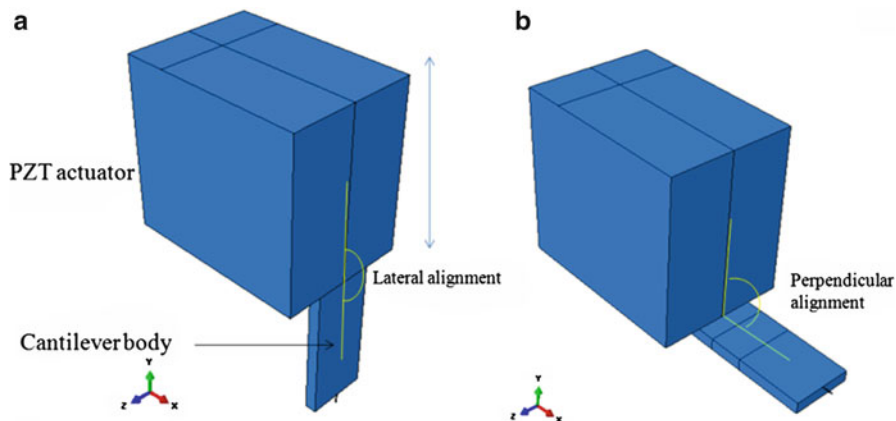
Fig. 19.5 (**a**). Assemble of the lateral configuration. (**b**). Assemble of the perpendicular configuration

On the other hand, for the perpendicular configuration of the nanoneedle, cantilever frame was attached with the PZT perpendicularly, i.e., angle between the PZT actuator and the cantilever is 90° (Fig. 19.5b). For both of the configurations, nanoneedle and PZT vibrate in *Y*-axis, but in the perpendicular configuration, cantilever body lays on *X*-axis and as a result, the cantilever remains hanging with less mechanical support than the lateral configuration. Figure 19.6 depicts the real-time image of the lateral configuration. Cantilever was attached with the PZT actuator using electric conductive tap and the inset shows the SEM image of the side view of the lateral nanoneedle. The purpose of different configurations is to analyze the steadiness of the nanoneedle when the position of the nanoneedle changes in design.

3   Calibration of the PZT-Actuated Nanoneedle for SCW Cutting

We have used both experimental and numerical methods to calibrate the assembled cell wall cutting device. The displacement of the attached fabricated nanoneedle was calibrated in terms of high input DC voltage. We have obtained high DC voltage by connecting three DC power supplies (*Agilent Technologies*) in series. Images of the nanoneedle were captured by IX73 inverted microscope, *Olympus Corp*.
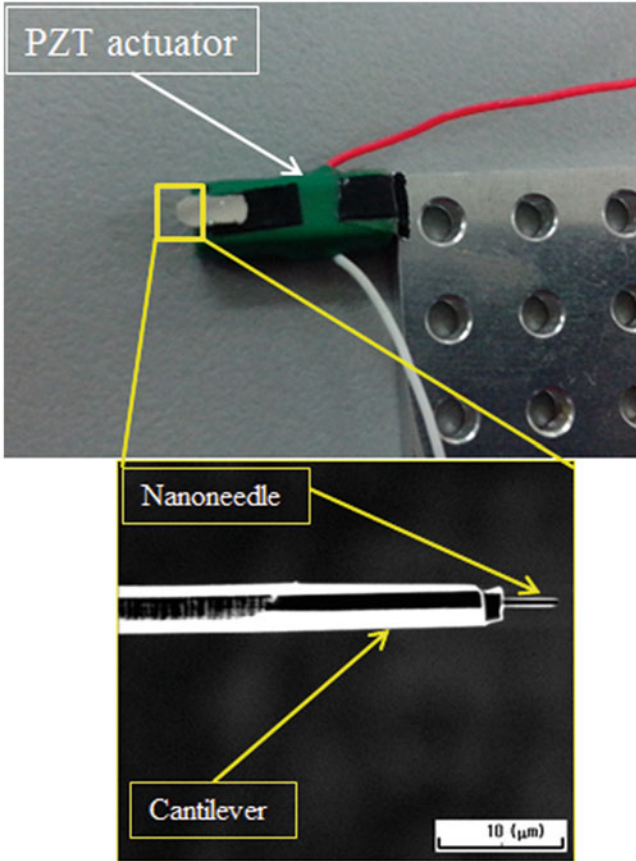
Fig. 19.6 Lateral configuration of the assembled device. *Inset* shows the SEM image of the lateral nanoneedle

3.1 Electrical Potential Analysis of the Actuator

In this research, we used the inverse piezoelectric effect to generate a partial displacement of the actuator in positive and negative *Y*-axis. Equation (19.1) describes the electric displacement in terms of applied voltage and piezoelectric matrix.

$$q_{i,j} = \varepsilon_{i,k,k} + D_{i,j,k}E_m \qquad (19.1)$$

where $q_{i,j}$ indicates electric charge displacement, $\varepsilon_{i,j,k}$ is the permittivity of the material, $D_{i,j,k}$ piezoelectric properties matrix, and E_m is the electric potential. These parameters are obtained from the datasheet of the PZT actuator and FE model was developed based on these parameters. Equation (19.2) was extracted from

the Maxwell's equation of motion and combined with Eq. (4.3) to illustrate the mechanical displacement (T). These equations were derived from the IEEE standard of piezoelectricity [23].

$$\Delta T = E_m . D_{i,j,k} \tag{19.2}$$

where ΔT indicates the mechanical displacement difference in terms of applied voltage for a piezoelectric actuator. Boundary conditions at the edge of the actuator were applied to obtain a very precise vibration of the actuator. If the material surface is discontinuous, then the boundary conditions across the surface are illustrated by Eqs. (19.4)–(19.6).

$$n_i T^I{}_{ij} = n_i T_{ij}{}^{II} \tag{19.3}$$

$$u_j{}^I = u_j{}^{II} \tag{19.4}$$

$$n_i q_i{}^I = n_i q_i{}^{II} \tag{19.5}$$

$$\varphi^I = \varphi^{II} \tag{19.6}$$

where I indicates the values of the variables on one side and II represents the other side of the surface. For traction-free surfaces, the boundary conditions are expressed in Eq. (19.7).

$$n_i T_{ij} = 0 \tag{19.7}$$

The surface without any displacement has the boundary condition Eqs. (19.8)–(19.10).

$$u_j = 0 \tag{19.8}$$

$$n_i q_i = 0 \tag{19.9}$$

$$\varphi = 0 \tag{19.10}$$

These boundary conditions were applied in X and Z directions of the actuator to constrain the vibration in $\pm Y$-axis only. Figure 19.7 showed the effect of applied voltage to the actuator. In our numerical analysis, we obtained a maximum displacement of 3.5 μm with an input voltage of 150 VDC.

a

U, Magnitude
+0.000e+00
+0.000e+00
+0.000e+00
+0.000e+00
+0.000e+00
+0.000e+00
+0.000e+00
+0.000e+00
+0.000e+00
+0.000e+00
+0.000e+00
+0.000e+00
+0.000e+00

At 0V

$\Delta T = 3.5 \mu m$

b

U, Magnitude
+3.484e-03
+3.194e-03
+2.903e-03
+2.613e-03
+2.323e-03
+2.032e-03
+1.742e-03
+1.452e-03
+1.161e-03
+8.710e-04
+5.807e-04
+2.903e-04
+0.000e+00

At 150V

Nanoneedle

Silicon Cantilever
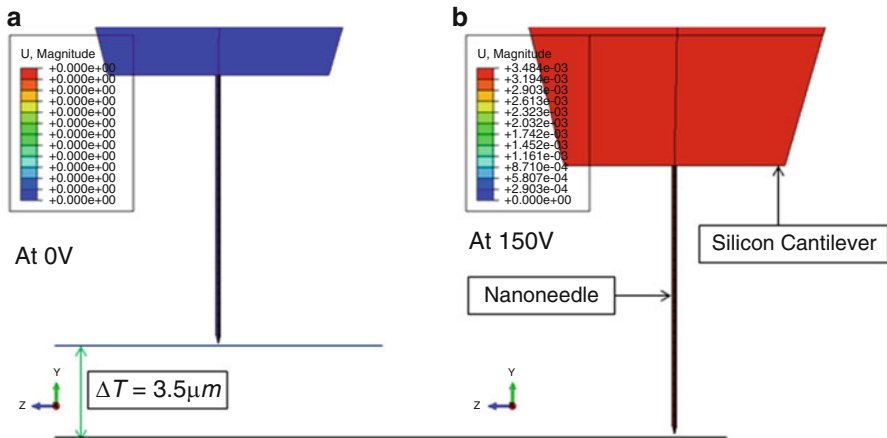
Fig. 19.7 Effect of electrical potential to the actuator. (**a**) Position of the nanoneedle at 0 V.
(**b**) Position of the nanoneedle at 150 V

3.2 Calibration of the Two Different Configurations of Assembling the Nanoneedle and PZT

In the previous section, it was showed that PZT actuator and nanoneedle were assembled in two different configurations, named as lateral and perpendicular configurations. In this section, we presented the calibration of each configuration in terms of applied voltage to the displacement.

3.3 Calibration of the Lateral Configuration

The input parameter of the actuator was electrical potential which controls the movement of the actuator. By adding three power supplies in series, we were able to manage a DC input voltage of 100 V, and at this input potential, the displacement of the nanoneedle was measured in real time. The electrical potential was increased from 0 to 100 V, at the same time the displacement of the nanoneedle tip was being measured. Figure 19.8a showed the position of the nanoneedle at 0 V and 19.8b at 100 V. A displacement of 4.5 μm was obtained from applied voltage of 100 V.

Actuator datasheet describes the displacement of the actuator is 4.5 ± 1.5 μm at maximum 150 VDC (AE0203D04F, NEC Corp.); suggested result is very much consistent with the obtained result i.e., 4.5 μm displacement. On the other hand, we have also performed the FE analysis of the modelled PZT actuator. Figure 19.9 shows the linear relationship between applied voltages with the nanoneedle displacement in *Y*-axis. The graph was plotted from a point node at the bottom surface of the actuator. Nodal elements were equally distributed to the entire
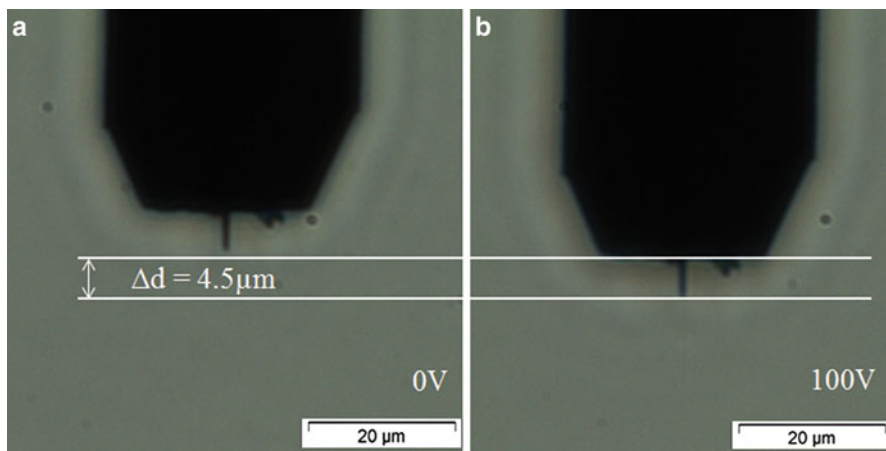
Fig. 19.8 Calibration of the displacement of rigid nanoneedle. (**a**). Nanoneedle position at 0 V and (**b**) at 100 V
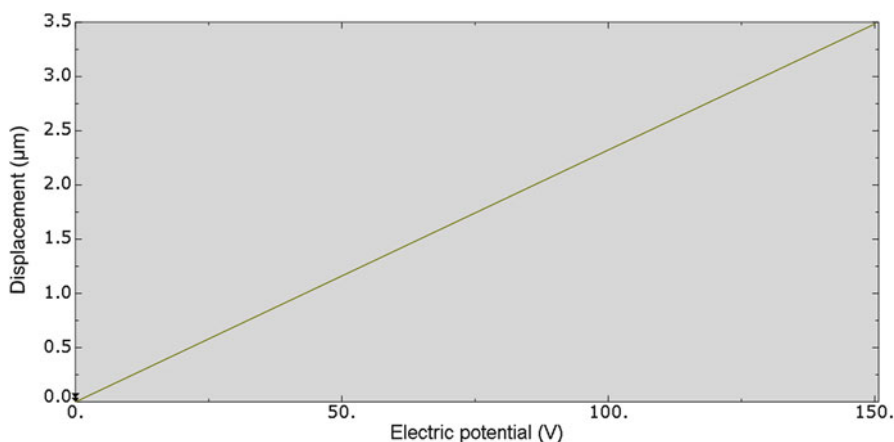
Fig. 19.9 Linear relationship between the applied electric potential and displacement of the nanoneedle

geometry. We measured the maximum displacement at 150 V towards Y-axis is 3.5 μm. Numerically obtained displacement profile is slightly different from the experimental result; the reason behind this is the inertia of the material and the discharging time. In FE model, ideal case of PZT actuator was considered. However, both of the experimental and numerical results are consistent with the datasheet of the product and stable for the application of the micro indentation in cell wall cutting.
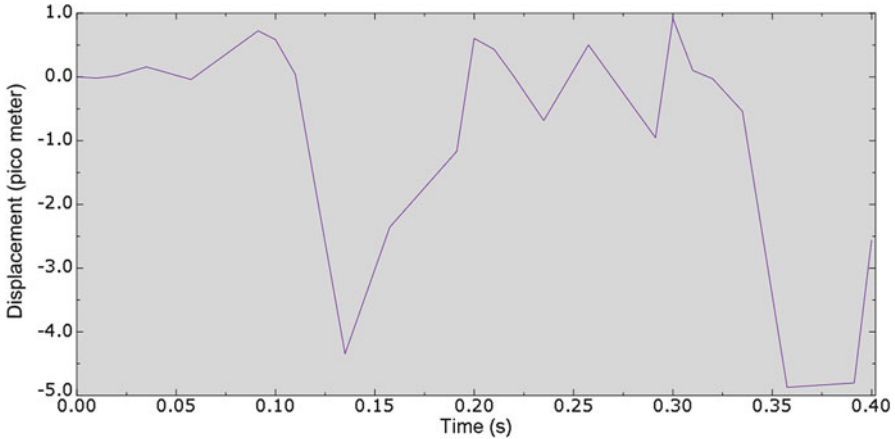
Fig. 19.10 Fractional displacement of the nanoneedle tip (for perpendicular configuration). Fractional displacement of the nanoneedle occurred only at the perpendicular configurations

3.4 Calibration of the Perpendicular Configuration

As the name implies, the position of the nanoneedle is perpendicular to the cantilever and attached with the PZT perpendicularly. Unlike in lateral configuration, we have further investigated the displacement of the nanoneedle in Z direction. Since the nanoneedle has been hung with the cantilever support only, the nanoneedle may have displacement in X or Z directions. Figure 19.9 showed the nanoneedle tip displacement in Y direction with input voltages of 150 V, which is same as the nanoneedle displacement as for the lateral configuration. Later on we observed the fractional displacement of the nanoneedle tip with time, at the same voltage. Figure 19.10 shows the spatial displacement of the nanoneedle tip in Z direction in variation of time.

After comparing the displacement profile of the lateral and perpendicular configurations, we observed that perpendicular configuration has a fractional displacement of the nanoneedle tip. At the lateral configuration, cantilever body and the piezoelectric actuator are placed 180°, i.e., both are in the same plane. On the other hand, at perpendicular configuration, the cantilever is in 90° angle to the actuator body, which caused the cantilever body to be hanging with the actuator body. Due to the length of the cantilever, there is some moment of inertia generated during the displacement of the nanoneedle and the cantilever body [24]. This fractional displacement behavior is not linear as the former configuration. Even though displacement is very small (1.2 pm), but in terms of cell wall cutting it has significant effect as it may cause cell wall damage. As a result, we conclude that lateral configuration is more stable than the perpendicular configuration.

4   Single Cell Wall Cutting Operations

4.1   FE Model of Yeast Cell

Saccharomyces cerevisiae W303 yeast cell was used as a sample model of a single cell. Yeast cells are eukaryotic microorganisms in the classification of fungi. The cell cycle of yeast is similar to human cell cycle [12] and yeast cell has been used as a sample cell for diagnosis of human disease such as tumor, cancer, and Parkinson's disease [25]. FE model of single yeast cell was developed based on the available mechanical properties of yeast cells. The geometric shape of the yeast cell is spherical and its diameter varies from 3 to 7 μm [12, 13]. The Young's modulus of yeast cell was reported as 3.21 MPa [13] and the Poisson's ratio of soft biological material was assumed as 0.5 [26]. In this work, the entire cell model was separated into two layers; cell wall and cell interior [27]. Figure 19.11 shows the model of yeast cell. Thickness of the yeast cell wall is about 100–300 nm [13] and the internal layer is cytoplasm which is the interior of the cell. Cytoplasm is a fluid-like structure with mass density of 1 g/mL [28]. Yeast cell wall is more likely to behave as a Mooney–Rivlin nonlinear hyperelastic material, as proposed by Stenson [29]. This model was extended by Lardner and Pujara [30], keeping the volume constant. Equation (19.11) illustrates the strain energy (*W*) function of the Mooney–Rivlin hyperelastic material used in the cell wall model.

$$W = C_1 (I_1 - 3) + C_2 (I_2 - 3) \tag{19.11}$$
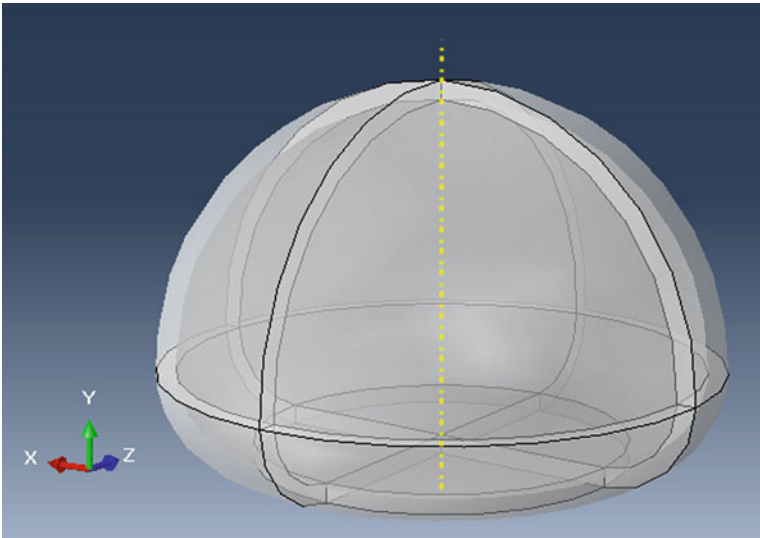
Fig. 19.11   Double layer model of yeast cell. Outer part is the cell membrane while the inner part is filled with cytoplasm (Eulerian materials)
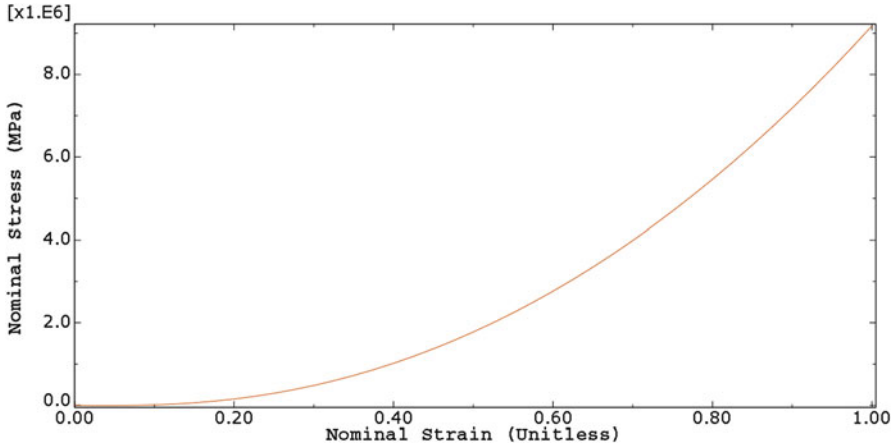
Fig. 19.12 Hyperelastic behavior of the yeast cell wall model

where C_1 and C_2, are the material constants with first and second invariant of I_1 and I_2, respectively. Nominal stress and strain of the yeast cell wall were obtained from [29] and further modified with [13]. The ultimate failure of cell wall occurs in 68 % of fractional deformations [31]. Figure 19.12 illustrated the material evaluation (stress vs. strain) of the nonlinear hyperelastic behavior of the yeast cell wall model. This model has a consistency with the proposed yeast cell wall model of Stenson [29], by which we could claim that developed yeast cell wall model is valid.

4.2 Effects of Vibration Frequency of Cell Wall Cutting

Since calibration results showed that lateral configuration is more stable in terms of displacement, we have used the lateral configuration of the nanoneedle to perform cell wall cutting operation. To investigate the effect of nanoneedle's vibration frequency to the cell wall damage, two different frequencies of the vibration were used; firstly at lower frequency (1 Hz) the penetration velocity was 7 μm/s. According to the piezoelectric theory, electric discharge occurred during compression (negative displacement). Several experiments were conducted to set the optimum frequency of the piezoelectric actuator. The maximum vibration frequency that we managed to vibrate was 15 Hz. Above that frequency, we were not able to observe the displacement under *IX73 Olympus microscopy*. As a result, for the highest frequency we chose to 10 Hz, where nanoneedle velocity was 24 μm/so. At both frequencies, the nanoneedle was able to cut the cell wall successfully. Figure 19.13 shows the depth of the damaged cell wall is 1.2 μm from its outer surface and cell wall thickness is 100–300 nm. This means, the cell wall was permanently penetrated during cutting operation. There were 12 cycles of penetration to the cell wall. We
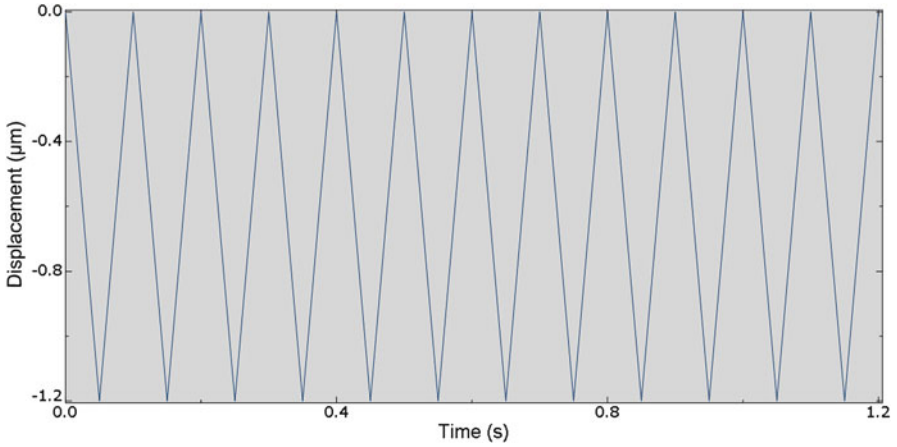
Fig. 19.13 Constant depth of the damaged (tip indentation) cell wall for 12 cycles at 1 Hz frequency

investigated the total wounded area during cell wall cutting in different frequencies. The total wounded area was calculated based on Eq. (19.12).

$$|d| = \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2} \qquad (19.12)$$

where d is the length of the cutting surface, x, y, z are the initial positions, and x_1, y_1, z_1 are the final positions of the nanoneedle. The coordinate points can be identified by inspecting the position of the nodes. From two nodes we can measure the distance easily. From Fig. 19.14a–d, we observed that the wounded area of the cell surface is not uniform. In order to avoid the complexity, we calculated the average wounded area of the cell surface. At 1 Hz (7 μm/s) frequency's nanoneedle we calculated the wounded length as 2.6 μm, width 1.07 μm, and the depth 1.2 μm. Surface area and volume were calculated by Eqs. (19.13) and (19.14), respectively.

$$A = l \times w \qquad (19.13)$$

$$V = l \times w \times h \qquad (19.14)$$

where l, w, h represents length, width, and height, respectively. From Eq. (19.13), the damaged surface area is calculated as 2.78 pm^2, and the volume is 3.39^{-18} m^3. Figure 19.15 shows the gradual profile SCW cutting for 10 Hz vibration of the nanoneedle. At 10 Hz frequency, the velocity of the nanoneedle was 24 μm/s and the wounded length was calculated as 2.31 μm, width 0.59 μm, and the depth 1.2 μm by nodal position. The volume of the wounded area was 1.63^{-18} m^3. For both of frequencies, SCW cutting operations were performed at different tip edge of the
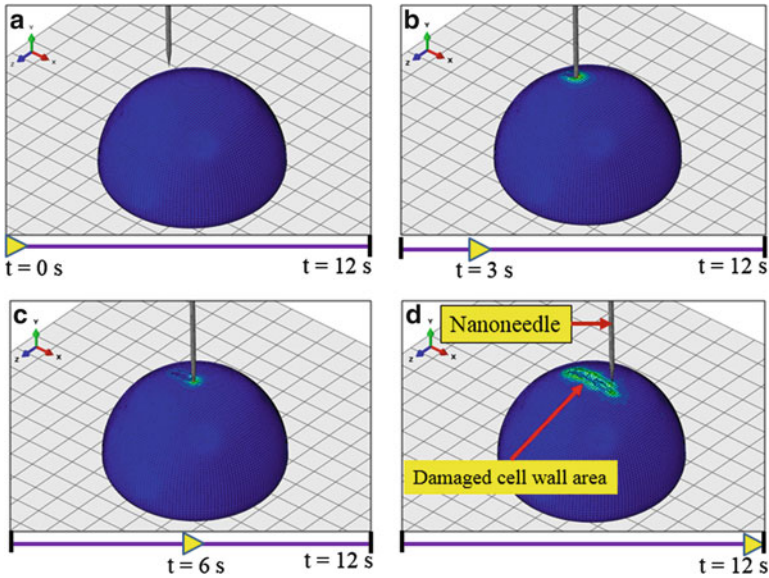
Fig. 19.14 Single yeast cell wall cutting operations at 1 Hz frequency. (**a**) At $t = 0$ s, just immediate before cutting operation. (**b**), (**c**) Single cell wall cutting operations at 3 s and 6 s, respectively. (**d**) Cell wall damaged area after cutting operations
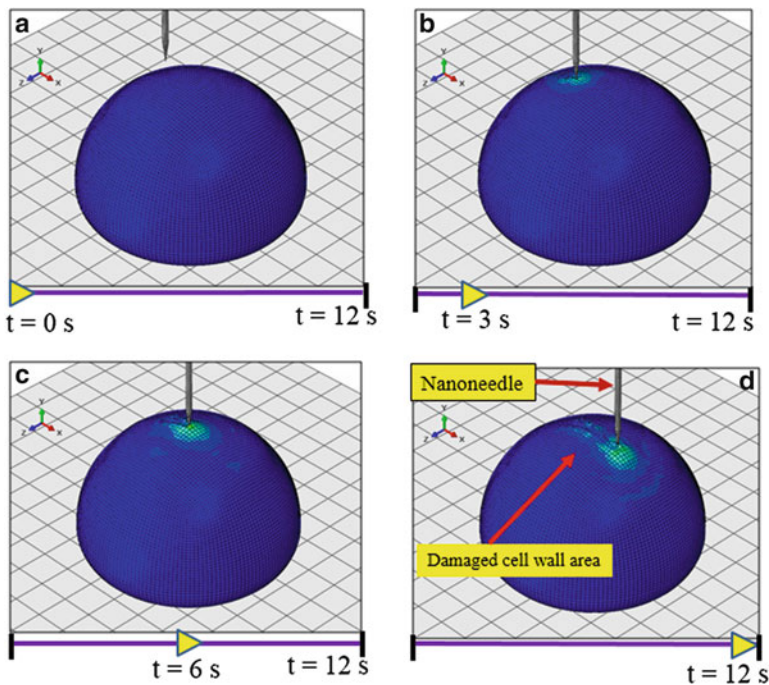
Fig. 19.15 Single cell wall cutting operations at 10 Hz vibrating frequency of the nanoneedle. Time lapse (**a–d**) illustrates the gradual cutting profile of the single cell wall

Table 19.2 Summary of the cell wall cutting with different edge angles of nanoneedle at constant indentation depth of 1.2 μm

Tip opening angle	Damaged area at 1 Hz frequency (pm^2)	Damaged area at 10 Hz frequency (pm^2)	Penetration force (nN)
10°	2	1.3	435
15°	2.78	2	616
20°	3.6	2.5	900

nanoneedle. Table 19.2 summarizes that for a constant nanoneedle edge angle, cell wall damaged area decreases with the increasing of the vibration frequency. For example, for 15° tip opening angle, cell wall wounded area is 2.78 pm^2 at 1 Hz frequency's nanoneedle while it is 2 pm^2 for 10 Hz frequency's nanoneedle, i.e., faster nanoneedle causes less damage to the cell wall. At both of the frequencies, cell wall penetration force remains unchanged for a constant penetration depth of 1.2 μm.

4.3   Effect of Nanoneedle Tip Edge Angle

During cell wall cutting, the nanoneedle was in a direct contact with the cell. Cell wall penetration force depends on the cell-tip contact area which varies with the changes of nanoneedle tip edge angle. The behavior of the cell-tip interaction can be described by the continuum mechanics of elastic contact. In extension to the Hertz–Sneddon model, the load exerted by the tip is linked to the indentation depth during cell wall cutting. If the geometrical parameters are known [26] for the conical nanoneedle tip, this behavior can be expressed by Eq. (19.15).

$$F_{cone} = \frac{2}{\pi} \tan \alpha \frac{E}{(1-v^2)} \delta^2 \tag{19.15}$$

where F is the load exerted by the tip, E the Young's modulus and v the Poisson's ratio of the cell wall, α opening edge angle of a conical tip and δ is the depth of the damaged cell wall [13]. Figure 19.16 showed the schematic of the tip edge angle for cell wall cutting. Penetration depth δ indicates the indentation of the nanoneedle inside the cell wall. From the calibration results, we came to know that PZT actuator is able to displace up to 4.5 μm, where 1.2 μm penetrated the cell wall and rest remain outside of the cell. For a 10° edge angle's nanoneedle, cell wall penetration force was 435 nN and penetration force increases with the increasing of nanoneedle edge angle. Previously Shen et al. reported single cell slicing required a force of 65 μN to make two slices. On the other hand, in this work cutting operations were performed on the cell wall, as a result force required is much lower than the cell slicing (maximum 900 nN). As SCW thickness is only 300–400 nm, it is confirmed that nanoneedle tip successfully penetrated single yeast cell. Higher penetration to the cell interior requires higher penetration force.
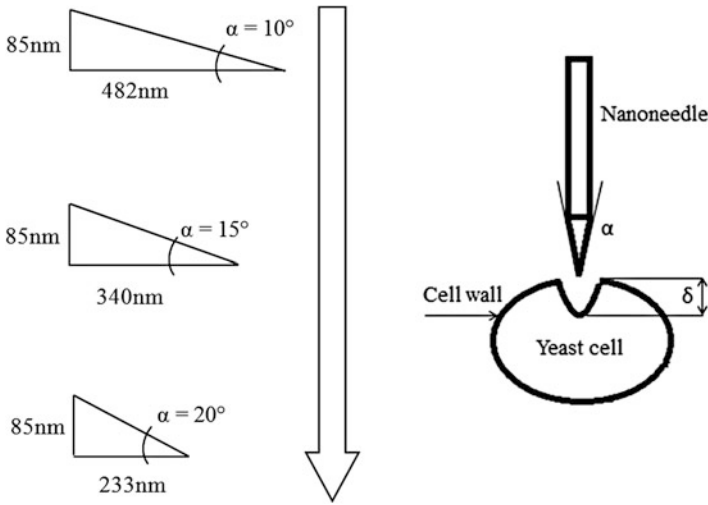
Fig. 19.16 Effect of nanoneedle edge angle on single cell wall cutting
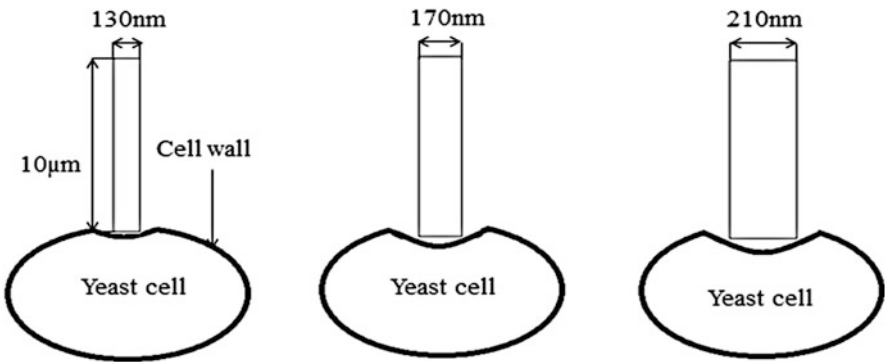
Fig. 19.17 Schematic diagram of the flat tip nanoneedle for cell wall cutting

4.4 Single Cell Wall Cutting with Flat Tip Nanoneedle

Previously we have reported, for a 316 nm diameter of flat nanoneedle tip, SCW penetration force is about 290 nN [13] inside environmental scanning electron microscopy (ESEM). But that work was limited to a particular diameter of the flat nanoneedle only. In this work, we have investigated the cell wall penetration force for possibly fabricated different diameters of the flat nanoneedle. Figure 19.17 shows the schematic diagram of the different diameters of flat nanoneedle. The area of the nanoneedle's tip increased with the diameter, which causes larger damage to the cell wall. In this case, the cell wall penetration forces were also increased.

Table 19.3 Cell wall penetration force at constant indentation depth of 1.2 μm

Diameter of nanoneedle (nm)	Penetration force (nN)
130	596
170	779
210	965

The Hertz–Sneddon relation between penetration force and indentation depth for a flat tip nanoneedle is illustrated in Eq. (19.16).

$$F_{cylindrical} = 2\frac{E}{(1-v^2)}a\delta \qquad (19.16)$$

where, E, a, δ, v are elastic modulus, radius, indentation, and Poisson's ratio, respectively. For a 130 nm diameter of flat nanoneedle, SCW penetration force is 596 nN. Table 19.3 summarizes the SCW penetration force for a constant nanoneedle indentation of 1.2 μm for different diameters of the cylindrical nanoneedle. Study was performed for three different sizes of the nanoneedle (130, 170, 210 nm). With the increasing of nanoneedle diameter, cell wall–nanoneedle contact area increases, which causes penetration force to be increased as well.

5 Conclusion

SCW cutting operation has been demonstrated successfully in this chapter. A piezoelectric-actuated rigid nanoneedle was used to control the movement of the nanoneedle. The nanoneedle and the PZT actuator were assembled in two different configurations: lateral and perpendicular configurations. We have calibrated the assembled device in terms of input potential and displacement. In terms of the nanoneedle tip displacement, lateral configuration is more stable than the perpendicular configuration. As a consequence, we performed cell wall cutting operation with the lateral configuration of the nanoneedle. Results showed that cell wall penetration force increases with the nanoneedle tip angle. We have also investigated the effect of flat tip nanoneedle to cell wall cutting. Larger diameter of the nanoneedle required a larger cell wall penetration force as tip-to-cell wall contact area increases with the increase of the diameter of the nanoneedle. In the future, we are going to perform SCW cutting of diseased cells and the results can be compared with a healthy human cell which will lead us to differentiate healthy and unhealthy cells based on the SCW strength.

References

1. Ahmad MR, Nakajima M, Kojima S et al (2010) Nanoindentation methods to measure viscoelastic properties of single cells using sharp, flat, and buckling tips inside ESEM. IEEE Trans Nanobiosci 9:12–23. doi:10.1109/TNB.2009.2034849

2. Zheng H, Chang L, Patel N et al (2008) Induction of abnormal proliferation by non-myelinating schwann cells triggers neurofibroma formation. Cancer Cell 13:117–128. doi:10.1016/j.ccr.2008.01.002

3. Beasley MB (2008) Pulmonary neuroendocrine tumours and proliferations: a review and update. Diagn Histopathol 14:465–473. doi:10.1016/j.mpdhp.2008.07.007

4. Assoian RK, Klein EA (2008) Growth control by intracellular tension and extracellular stiffness. Trends Cell Biol 18:347–352. doi:10.1016/j.tcb.2008.05.002

5. Taddei ML, Giannoni E, Comito G, Chiarugi P (2013) Microenvironment and tumor cell plasticity: an easy way out. Cancer Lett. doi:10.1016/j.canlet.2013.01.042

6. Amako K, Takade A, Umeda A, Yoshida M (1993) Imaging of the surface structures of epon thin sections created with a glass knife and a diamond knife by the atomic force microscope. J Electron Microsc (Tokyo) 42:121–123

7. Chang CC, Chen PH, Chu HL et al (2008) Laser induced popcornlike conformational transition of nanodiamond as a nanoknife. Appl Phys Lett 93:33905. doi:10.1063/1.2955840

8. Ahmad MR, Nakajima M, Kojima S et al (2008) In situ single cell mechanics characterization of yeast cells using nanoneedles inside environmental SEM. IEEE Trans Nanotechnol 7:607–616

9. Singh G, Rice P, Mahajan RL, McIntosh JR (2009) Fabrication and characterization of a carbon nanotube-based nanoknife. Nanotechnology 20:095701. doi:10.1088/0957-4484/20/9/095701

10. Ceusters J, Londers E, Brijs K et al (2008) Glucuronoarabinoxylan structure in the walls of Aechmea leaf chlorenchyma cells is related to wall strength. Phytochemistry 69:2307–2311. doi:10.1016/j.phytochem.2008.06.002

11. Fineran BA (1994) Hot fixation of fungal spores for transmission electron microscopy: application to thick-walled spores of the smut fungus Entorrhiza. Mycol Res 98:799–809. doi:10.1016/S0953-7562(09)81059-9

12. Shen Y, Nakajima M, Yang Z et al (2011) Design and characterization of nanoknife with buffering beam for in situ single-cell cutting. Nanotechnology 22:305701. doi:10.1088/0957-4484/22/30/305701

13. Ahmad MR, Nakajima M, Kojima S et al (2008) The effects of cell sizes, environmental conditions, and growth phases on the strength of individual W303 yeast cells inside ESEM. IEEE Trans Nanobiosci 7:185–193. doi:10.1109/TNB.2008.2002281

14. Cheng CH, Tu SL (2008) Fabrication of a novel piezoelectric actuator with high load-bearing capability. Sensors Actuators A Phys 141:160–165. doi:10.1016/j.sna.2007.07.033

15. Ronkanen P, Kallio P, Vilkko M et al (2011) Displacement control of piezoelectric actuators using current and voltage. IEEE Trans Mechatronics 16:160–166

16. Rahman MH, Sulaiman AH, Ahmad MR, Fukuda T (2013) Finite element analysis of single cell wall cutting by piezoelectric-actuated vibrating rigid nanoneedle. IEEE Trans Nanotechnol 12:1158–1168

17. Schnabel JA, Tanner C, Castellano-Smith AD et al (2003) Validation of nonrigid image registration using finite-element methods: application to breast MR images. IEEE Trans Med Imaging 22:238–247. doi:10.1109/TMI.2002.808367

18. Rathore JS, Majumdar R, Sharma NN (2012) Planar wave propagation through a tapered flagellated nanoswimmer. IEEE Trans Nanotechnol 11:1117–1121. doi:10.1109/TNANO.2012.2214230

19. Salmanogli A, Rostami A (2012) Modeling and improvement of breast cancer site temperature profile by implantation of onion-like in tumor site. IEEE Trans Nanotechnol 11:1183–1191

20. Škoro GP, Bennett JRJ, Edgecock TR et al (2011) Dynamic Young's moduli of tungsten and tantalum at high temperature and stress. J Nucl Mater 409:40–46. doi:10.1016/j.jnucmat.2010.12.222
21. Lazarus A, Thomas O, Deü JF (2012) Finite element reduced order models for nonlinear vibrations of piezoelectric layered beams with applications to NEMS. Finite Elem Anal Des 49:35–51. doi:10.1016/j.finel.2011.08.019
22. Kanda T, Kurosawa MK, Higuchi T (2002) Sensitivity of a miniaturized touch probe sensor using PZT thin film vibrator. Ultrasonics 40:61–65
23. (1988) An American National Standard IEEE Standard on Piezoelectricity
24. Wang Z, Zhu W, Zhao C, Yao X (2001) Deflection characteristics of a trapezoidal multi-layer in-plane bending piezoelectric actuator. IEEE Trans Ultrason Ferroelectr Freq Control 48:1103–1110
25. Lehle L, Strahl S, Tanner W (2006) Protein glycosylation, conserved from yeast to man: a model organism helps elucidate congenital human diseases. Angew Chem Int Ed Engl 45:6802–6818. doi:10.1002/anie.200601645
26. Svaldo Lanero T, Cavalleri O, Krol S et al (2006) Mechanical properties of single living cells encapsulated in polyelectrolyte matrixes. J Biotechnol 124:723–731. doi:10.1016/j.jbiotec.2006.02.016
27. Sulaiman AH (2014) Integrated dual nanoprobe-microfluidics system for single cell electrcial property characterizations
28. Bryan AK, Goranov A, Amon A, Manalis SR (2010) Measurement of mass, density, and volume during the cell cycle of yeast. Proc Natl Acad Sci 107:999–1004. doi:10.1073/pnas.0901851107
29. Stenson JD, Hartley P, Wang C, Thomas CR (2011) Determining the mechanical properties of yeast cell walls. Biotechnol Prog 27:505–512. doi:10.1002/btpr.554
30. Lardner TJ, Pujara P (2008) Compression of spherical cells. In: Nasser SN (ed) Mechanics today, vol 5. Pergamon Press, New York, pp 161–176
31. Smith A, Zhang Z (2000) The mechanical properties of Saccharomyces cerevisiae. Proc Natl Acad Sci 97:2–5

Chapter 20
A Robotic Percussive Riveting System for Aircraft Assembly Automation

Fengfeng (Jeff) Xi, Yu Lin, and Yuwen Li

Abstract Presented in this chapter is a robotic percussive riveting system for aircraft assembly automation. It is shown here that a successful robot application to the automation of a process requires in-depth research of the process and interaction with the robot. For this purpose, a process planning-driven approach is proposed to guide this development. A typical process planning will involve a list of key considerations including process sequence, process parameters, process tooling, and process control. Through this list, a number of key issues are identified for the robotic percussive riveting process, such as rivet pattern planning, rivet time determination, rivet tooling design, and rivet insertion control. Furthermore, an important issue pertinent to robot interaction is identified, i.e., robot fatigue life under repetitive percussion during riveting. It is demonstrated here that the thorough research on these issues has effectively created know-how for the successful development of our robotic percussive riveting system.

Keywords Robotic riveting • Percussive riveting • Robotic assembly • Aircraft assembly robot • Robotic process planning • Percussive riveting process modeling • Robotic riveting tooling design • Rivet insertion visual servoing control • Robot fatigue analysis

1   Introduction

Riveting and welding represent two primary joining methods for the assembly of structural components that require strong joint strength. Compared with welding mainly a fusion method, riveting a mechanical method generates no thermal

F. (Jeff) Xi (✉)
Department of Aerospace Engineering, Ryerson University, Toronto, ON, Canada
e-mail: fengxi@ryerson.ca

Y. Lin
Kirchhoff Van-Rob, Aurora, ON, Canada

Y. Li
School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China

deformation, hence widely used in aircraft assembly for joining high thermal conductive materials such as aluminum sheet metals [1]. There are hundred thousands of rivets in a regional aircraft and millions in a large continental aircraft. Overall, the operation of aircraft assembly is divided into three stages: subcomponent assembly, component assembly, and line assembly. The subcomponent assembly is the first step to construct the base components for four major sections, namely, fuselage, wing, cockpit, and empennage. The component assembly is the middle step to join the subcomponents to form an individual major section. The line assembly is the last step to assemble a whole aircraft by connecting the four major sections together.

The current riveting process in aerospace manufacturing entails a mix of manual riveting, semi-automated riveting, and automated riveting. The semi-automated and automated riveting machines are widely used in North America and Europe, but only limited to component assembly, such as wing skin panels and fuselage skin panels. Subcomponent assembly and line assembly are still conducted manually. The labor incurred producing these subassemblies/assemblies amounts to as much as 50 % of the total cost. Manual riveting operations are tedious, repetitive, and prone to error, likely causing health and ergonomic problems [1].

In principle, there are two riveting methods, the first called *squeezing* (or one-shot) riveting, where a large upsetting force is applied to deform a rivet instantly. This method requires a large riveter operating under high pressure beyond the yield strength of aluminum rivets in a range over 500 lb force. As shown in Fig. 20.1a, this type of riveter is made of either a hydraulic cylinder or an electromagnetic piston, very heavy, bulky, and usually needing a lift-assisted device if used for manual operation. The automated and semi-automated riveting machines employ this type of riveter; hence they are gigantic and only limited to riveting large, simple, and relatively flat components. The second method is called *percussive* (or hammering) riveting, where a small impulsive force is applied to deform a rivet accumulatively by a series of hits. As shown in Fig. 20.1b, this method uses a rivet gun in size of a regular handheld power tool, very compact and light, operating under much lower pressure in a range less than 100 psi, very safe and energy efficient. Manual riveting employs this principle.

Research on robotic riveting has been mainly centering on squeezing riveting that utilizes heavy-duty industrial robots of large size (>100 Kg payload). In the automotive industry, squeezing robotic riveting systems have been fully developed and commercialized for joining metal parts. This technology is called robotic self-piercing riveting, in which a C-frame tooling, as shown in Fig. 20.1c, is designed to mount a squeezing riveter on one end as a punch and the other end serving as a hitting base [2]. This system has been widely used for automotive chassis assembly. The application of robotic technology in aerospace manufacturing has been significantly slower than that in automotive manufacturing [3]. Though not commercially available, squeezing robotic riveting systems have been researched in the past by Boeing [4] and recently by EADS in Germany affiliated with AirBus [5]. In addition, a robotic system has been implemented at Bombardier in Montreal that uses two giant Kuka robots to hold large panels that are riveted on a C-frame squeezing riveting machine [6].
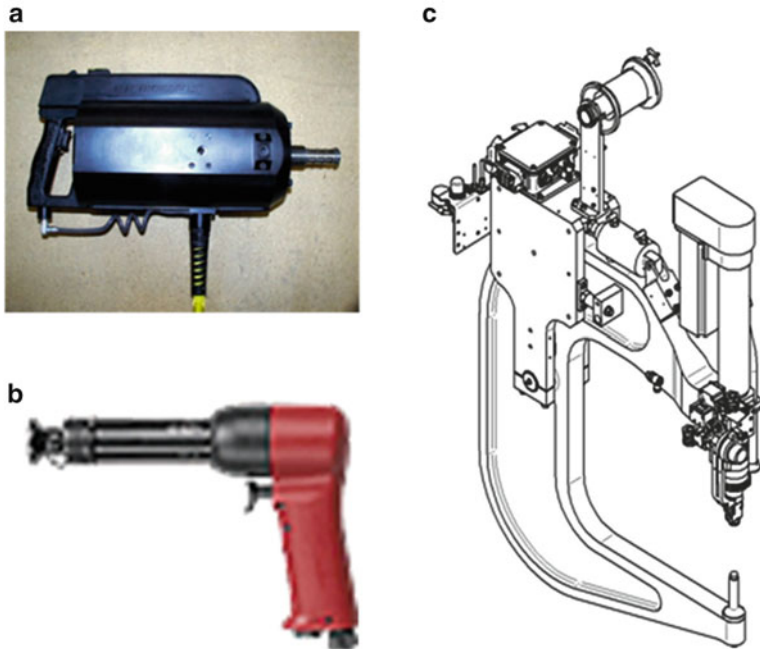
Fig. 20.1 Various rivet tools. (**a**) Squeezing riveter: length >24", weight > 50 lbs. (**b**) Percussive rivet gun: length <10" weight < 5 lbs. (**c**) C-frame rivet tooling

Though adhesives are used to bond composites, riveting remains as a primary method for joining composite panels where there is requirement for strong joint strength and prevention of laminate de-bonding. Automated squeezing riveting systems have been developed by AirBus and Boeing for riveting composite panels of fuselages and wings. As composites are being introduced to replace steels for fabrication of automotive structural parts, robotic riveting will likely take over welding as a primary joining method for the future of automotive industry [7].

By comparison, robotic percussive riveting is much more compact. Not only a much smaller riveting gun is used but also a light-/medium-payload industrial robot of small size (<50 Kg payload) can be applied. The overall system compactness offers a great advantage that a robotic percussive riveting system is able to access tight and awkward areas that a squeezing robotic riveting system is not able to. This advantage is referred to as good tool accessibility. In this chapter a framework of research is presented to show that the successful implementation of a robot application requires in-depth process research pertinent to the application.

2 Framework Overview

The initial application of industrial robots was concentrated on pick-and-place applications with grippers as the main tooling. Since then, they have been explored for various manufacturing applications both contact and non-contact. A good example of noncontact robot application is robotic welding, widely practiced in the automotive industry, where a weld gun is used not in contact with the workpiece. A good example of contact robot application is robotic polishing, widely adopted in the automotive and aerospace industry, where a polishing tool is used in contact with the workpiece. The robot companies who sell robots for these applications usually offer add-on modules, e.g., robot motion planning software. This indicates that the successful implementation of a robot application requires understanding of the process itself and interaction with the robot.

For this reason, we have developed a process planning-driven approach for the development of our robotic percussive riveting system, a new application for aircraft assembly automation. Figure 20.2 shows the developed robotic percussive riveting system. It includes a six-DOF industrial robot that replaces the first worker for holding/moving a percussive rivet gun and a five-axis computer numerical control (CNC) gantry system that replaces the second worker for holding a bucking bar. The entire riveting process is automated through synchronization between the robot and gantry. Furthermore, the choice of a gantry system instead of a second robot allows it to serve as a jig for mounting sheet metals.
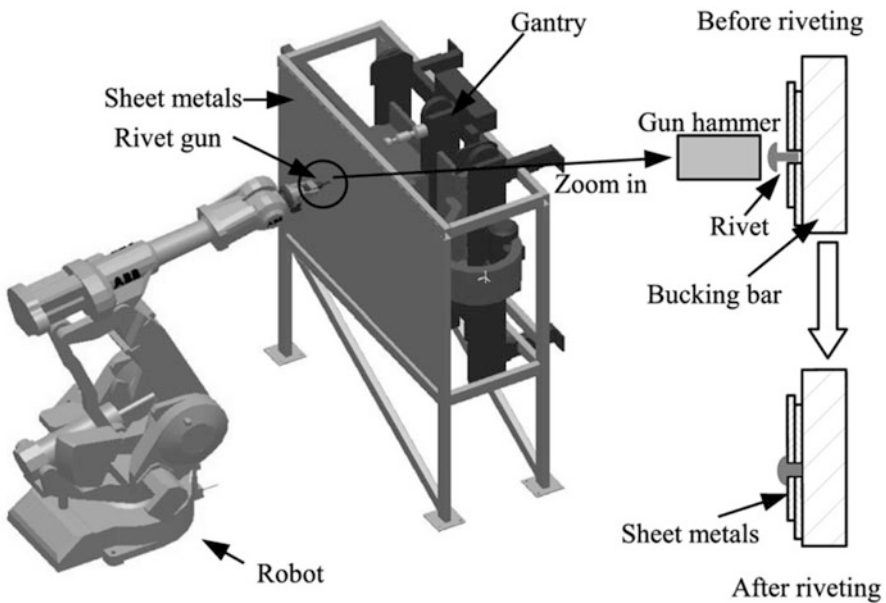
Fig. 20.2 Developed robotic percussive riveting system [13]

Table 20.1 Process-planning driven research

Process planning	Required research
Process sequence	Riveting pattern planning
Process parameters	Riveting time determination
Process tooling	Rivet tooling design
Process control	Rivet insertion control

Given that the goal of process planning is to generate a plan in order to successfully produce a product [8], this planning contains a number of key considerations: *process sequence*, *process parameters*, *process tooling*, and *process control*. As outlined in Table 20.1, these key considerations are used to guide the development of our robotic percussive riveting system. In what follows, our research methods are described.

3   Process Planning-Driven Approach

3.1   Process Sequence

Process sequence deals with the steps of a process, which is process specific. Commercial robot planning programs do not provide this feature. Hence, there is a need to study the riveting process. Riveting requires drilling a hole first and then inserting a rivet for fastening. Two riveting processes are exercised in practice, *simultaneous* and *sequential*. The first one is to drill and rivet together on a rivet spot, which demands a large tooling to combine a drill and a rivet gun. This process is typically applied on large automated riveting machines where tool accessibility is not of concern for the assembly of large and flat panels. The second process is to drill a series of holes first at the required rivet spots and then switch to a rivet gun for riveting. This process keeps the tooling compact and light, with good tool accessibility for tight and awkward areas, and hence is employed for manual operation. Since our system is developed to replace manual operation, the second process is considered here.

In riveting, the required rivet spots are determined by rivet patterns in light of industry standards. For automation, there are two ways to obtain the coordinates of these spots. The first one is to extract from CAD models, as modern aircraft components are designed using CAD. The second way is to compute these spots directly according to rivet patterns. While our development can accommodate both, only the second case is presented here, as the first case is straightforward.

Figure 20.3 provides a taxonomy of rivet patterns. As depicted in Fig. 20.4, lap joints are formed by overlapping two pieces of sheet metal, which is asymmetric causing the secondary bending [9]. Butt joints, on the other hand, are created by aligning two pieces of sheet metal to maintain symmetry. Both joints can be laid out in single, double, and multiple rows as depicted in Fig. 20.5. Rivet patterns can be of chain type with rows lined up forming a grid or of zigzag type with rows offset up or down.
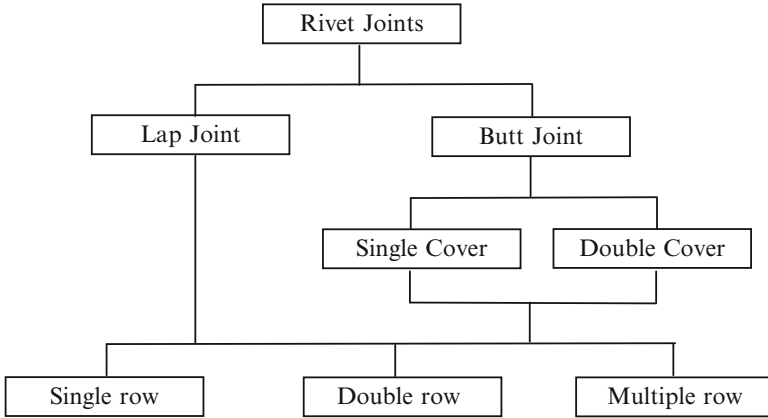
Fig. 20.3 Taxonomy of rivet patterns
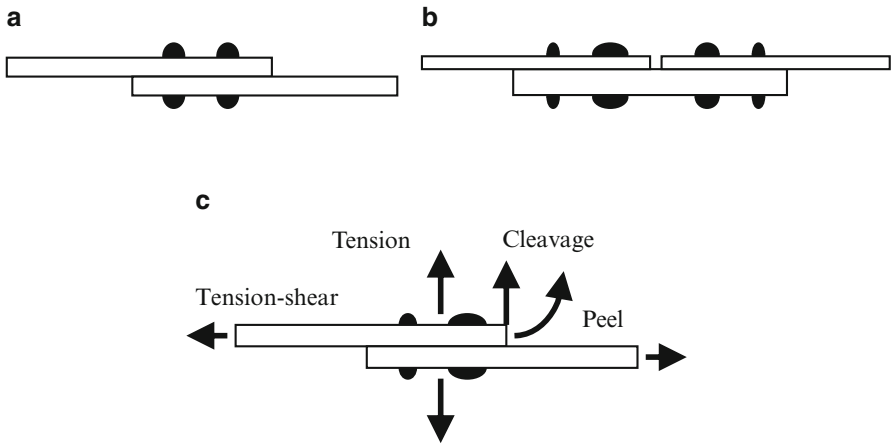
Fig. 20.4 Riveted joints. (**a**) lap joint, (**b**) butt joint (single cover), (**c**) joint strength
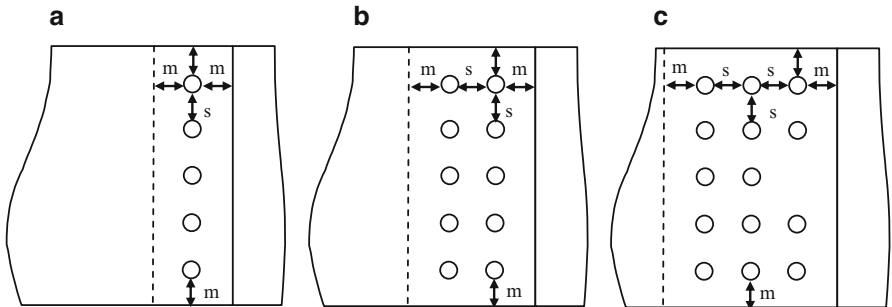
Fig. 20.5 Rivet rows. (**a**) single row, (**b**) double row, (**c**) multiple row

According to the aerospace standards [10], rivet size should be determined first with respect to the thickness of sheet metals, followed by rivet spacing. Symbolically, rivet size can be expressed as

$$[d, l] = f_1 (t_1, \ t_2) \qquad (20.1)$$

where d and l are the diameter and length of a rivet, respectively; t_1 and t_2 are the thickness of two pieces of sheet metal joining together. The computed rivet size must be rounded up to match with that specified by the standards, such as AN (Air-force Navy), MS (Military Standard), and NAS (National Aircraft Standard) [10].

Rivet spacing is a function of the rivet diameter and the riveting pattern, expressed as

$$[s, m] = f_2 (d, n) \qquad (20.2)$$

where s and m are the spacing between adjacent rivets and the margin on all sides, respectively; n denotes the number of rows. For given width of sheet metals, the number of rivets is determined from s and m. Aircraft rivets include solid rivets (requiring bucking bar) and cheery rivets (no bucking bars). Clecos are used as temporary fasteners. In this study, only solid rivets are considered.

A rivet planning software package has been developed that can compute all the rivet spots based on the abovementioned information. Figure 20.6 displays two snapshots of the software developed using ABB RobotStudio, with the first showing the animation window and the second the rivet spot planning window. Furthermore, this package is being developed to include joint strength analysis. Joint strength is defined as the joint's ability to resist against tension, shear, cleavage, and peel, as shown in Fig. 20.4c. In general, the riveted joint strength is proportional to the number of rows, called joint efficiency [11]. The common failures of the riveted joints are caused by the in-plane force (tension-shear shown in Fig. 20.4), including breaking of the sheet at the hole section, shearing of the rivet, crushing of the sheet and rivet, and shearing of the hole [10]. Joint strength is also pertinent to
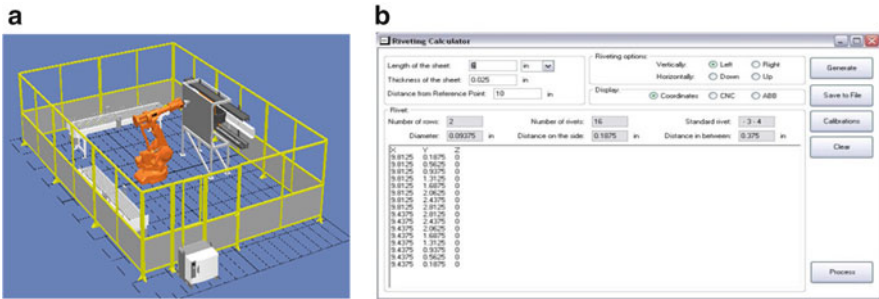
Fig. 20.6 Developed rivet planning software

the type of joints, symmetric better than asymmetric. Joint strength analysis can be performed using standard stress analysis methods [11] or more advanced finite element methods [12].

3.2 Process Parameters

As shown in Fig. 20.2, in percussive riveting a rivet is placed between a rivet gun and a bucking bar to subject to repetitive impulses from the hammer of the gun. Due to these impacts, the rivet is deformed plastically to join two pieces of sheet metal together. Upon the determination of riveting process sequence, both the robot holding the gun and the gantry holding the bucking bar can be programmed to follow a path specified according to a given rivet pattern and move from spot to spot. However, this program does not know how much time is needed to perform riveting at each spot. Hence, there is a need to study process parameters, which involves riveting process modeling.

The said modeling comprises two theories, impact dynamics and plasticity [13]. Impact dynamics is applied to model the kinetic energy generated by the percussive gun, and plasticity is applied to model the rivet plastic deformation caused by the impact. As shown in Fig. 20.7a, a percussive rivet gun is pneumatic driven and composed of a piston and a hammer. Under a compressed air supply the piston is pushed to drive the hammer to hit the rivet. As illustrated in Fig. 20.7b, at the start point of the piston stroke, the air pressure on its rear end is higher than that on its front end, so the piston moves forward. As it moves close to the end of its stroke, the pressure difference on the two ends reverses, thereby bringing the piston back. The stroke cycle repeats till the air supply is turned off. The reciprocal of the stroke cycle time is called the triggering frequency. Figure 20.8 shows a test result of the vibration experiment conducted to establish an empirical relation between the triggering frequency and the supply air pressure.

The key in impact dynamics modeling is to determine the hammer velocity hitting the rivet. First, the piston dynamics can be related to the air supply as, without consideration of friction

$$m_p a_p = Ap \qquad (20.3)$$

where m_p, a_p, A, and p represent the mass, acceleration, cross-section area, and air pressure of the piston, respectively. If the impact between the hammer and the piston is assumed perfectly elastic, the total momentum and total energy are conserved, hence

$$m_p \left(v_{po} - v_p \right) = m_h \left(v_h - v_{ho} \right) ; \quad v_p + v_{po} = v_h + v_{ho} ; \qquad (20.4)$$

where v_{po}, v_p, v_{ho}, and v_h represent the velocity of the piston and hammer before and after the impact, respectively; m_h denotes the mass of the hammer. Note that v_{po} is
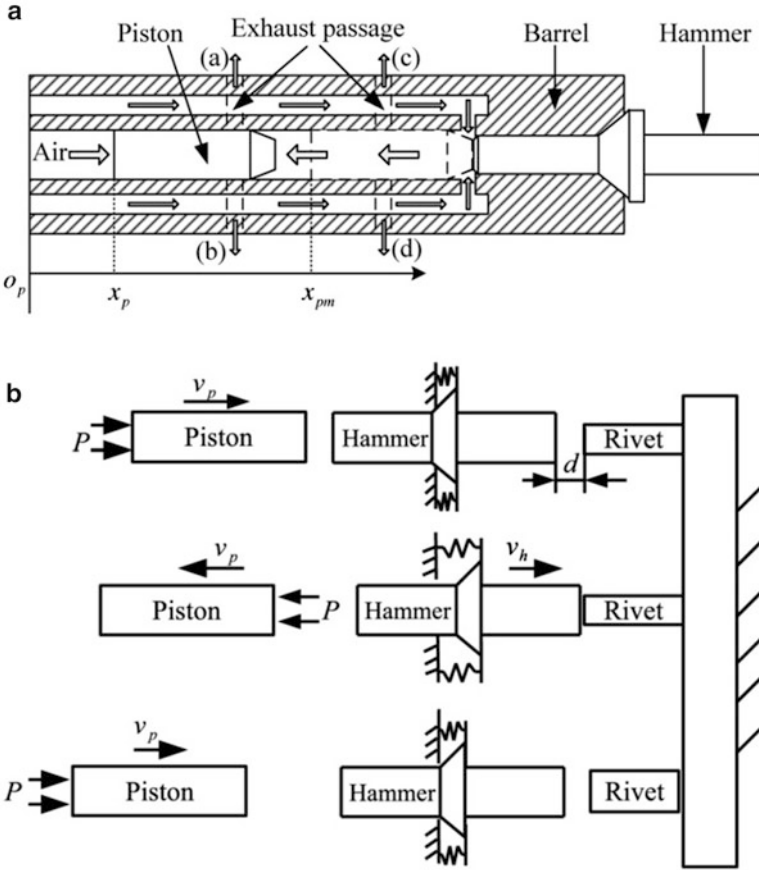
Fig. 20.7 Modeling of a percussive gun [13]. (**a**) Schematics of a percussive rivet gun. (**b**) Impact modeling of a percussive rivet gun

computed from Eq. (20.3) and v_{ho} is zero since the hammer is initially stationary; the hammer velocity can be derived from Eq. (20.4) as

$$v_h = \frac{2m_p v_{po}}{m_p + m_h} \tag{20.5}$$

The hammer velocity is the velocity hitting the rivet which in turns hits the bucking bar. Each hit induces a small rivet plastic deformation, and the accumulation of a series of hits results in a large rivet deformation. For this reason, the rivet is discretized into N elements, each modeled as a spring-mass-damper system, as shown in Fig. 20.9. The spring forces are modeled by a bilinear stress–strain curve containing both elastic and plastic deformation. Consequently, the dynamics
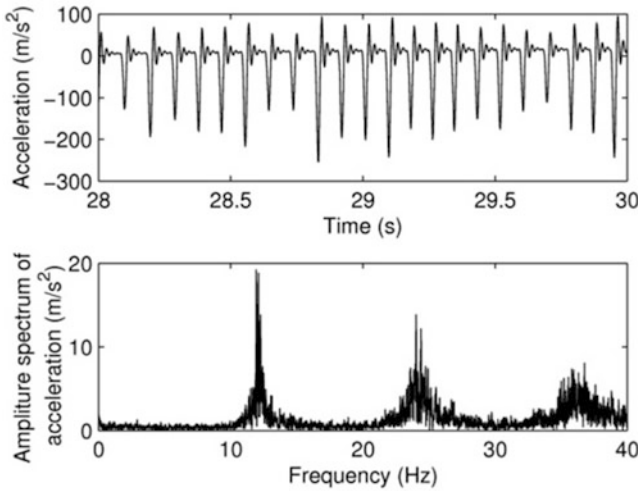
Fig. 20.8 Vibration experiment on percussive rivet gun [13]
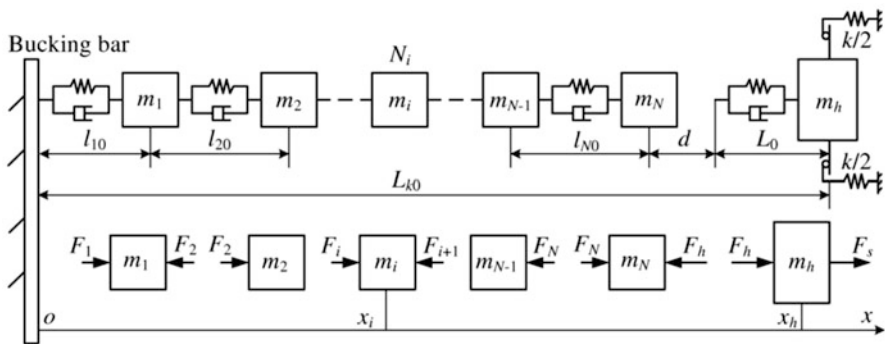
Fig. 20.9 Dynamic modeling of hammer, rivet, and bucking bar [13]

comprising the hammer, rivet, and bucking bar can be expressed by a set of $2N + 2$ first-order nonlinear ordinary differential equations as [13]

$$\dot{y} = F(y) \tag{20.6}$$

where y is a vector representing the displacements of the N elements and the hammer. Note that the bucking bar is fixed, no displacement. Equation (20.6) can only be solved by computer. The simulation result given in Fig. 20.10 illustrates how a rivet deforms incrementally under a series of hits.

This simulation program is being embedded into the rivet planning software shown in Fig. 20.6 for the purpose of computing the required rivet time at each rivet spot. To do so, first the required rivet deformation is obtained by calculating the difference between the original rivet length and the thickness of two pieces of
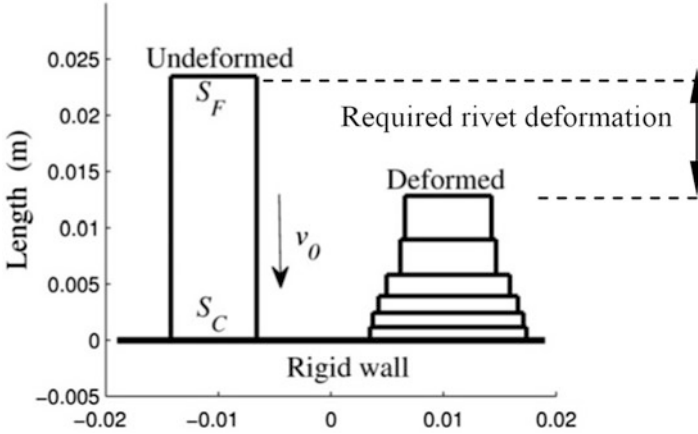
Fig. 20.10 Simulation of rivet plastic deformation [13]

sheet metal. Then, the simulation program will run to determine the number of hits needed to produce the required rivet deformation. At last, the required rivet time can be decided by multiplying the number of hits and the time interval of hits (i.e., reciprocal of the triggering frequency).

3.3   Process Tooling

Though with rivet path planned and rivet time determined, actual implementation still cannot be guaranteed unless the tool is ensured to do the job. Different riveting methods bring up different issues in tooling design. In the conventional squeezing riveting, a large static force is applied, the main concern being the robot rigidity to withstand the static force. In percussive riveting, however, a series of impulsive (relatively small) forces is applied; the main concern becomes robot vibration. The general guidance of robot tooling design states that the tool should be designed lightweight, in compact size, and with large holding force against vibrations [14]. In other words, the key issue is how to keep the tool small yet strong. Hence, there is a need to study tool design.

Attachment of a tool to the robot end-effector will change the system kinematics and dynamics. With the tool mounted, kinematic analysis should be carried out with respect to TCP (tool center point) instead of the center point of the end-effector (usually the center point of the mounting plate for the industrial robot). This analysis can be readily accommodated by treating the tooling system as an add-on body in the multi-body system of the robot, as shown in Fig. 20.11. Therefore, the system dynamic equations can be given as [15]

$$\left(\mathbf{M}_r + \mathbf{M}_t\right)\ddot{\mathbf{q}} + \left(\mathbf{C}_r\left(\mathbf{q}, \dot{\mathbf{q}}\right) + \mathbf{C}_t\left(\mathbf{q}, \dot{\mathbf{q}}\right)\right) + \left(\mathbf{G}_r\left(\mathbf{q}\right) + \mathbf{G}_t\left(\mathbf{q}\right)\right) = \boldsymbol{\tau} - \mathbf{J}_t^{\mathrm{T}}\mathbf{w}_t \quad (20.7)$$
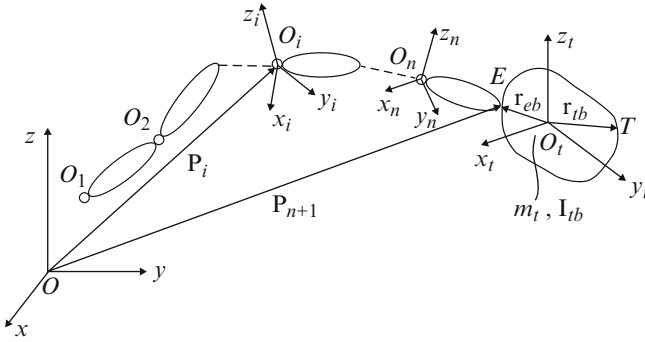
Fig. 20.11 Robot system with tooling [15]

where $\mathbf{q}$ is a vector of the robot joint displacements; $\mathbf{M_r}$, $\mathbf{M_t}$, $\mathbf{C_r}$, $\mathbf{C_t}$, $\mathbf{G_r}$, and $\mathbf{G_t}$ represent the matrices of robot mass, tooling mass, robot coupling term, tooling coupling term, robot gravitational term, and tooling gravitational term, respectively; $\boldsymbol{\tau}$ is a vector of joint actuation forces; $\mathbf{J_t}$ is the Jacobian of the TCP; $\mathbf{w_t}$ is the vector of wrench (force and moment) at TCP. Without the tool mounted, Eq. (20.7) would be reduced by setting $\mathbf{M_t}$, $\mathbf{C_t}$, and $\mathbf{G_t}$ to zero, and $\mathbf{J_t}$ and $\mathbf{w_t}$ to $\mathbf{J_r}$ and $\mathbf{w_r}$ of the robot, respectively. Apparently, the tooling system will affect the entire system.

Efforts have been made to relate the system dynamics to the aforementioned three considerations of the general guidance, thereby creating a new theory for the tooling design. For this purpose, three new indices are introduced to evaluate the influence of the tooling system on the overall robot system dynamics [15]. The first one, η, is a kinetic energy ratio defined as

$$\eta = \frac{\dot{q}^{\mathrm{T}} M_{\mathrm{r}} \dot{q}}{\dot{q}^{\mathrm{T}} (M_{\mathrm{r}} + M_{\mathrm{t}}) \dot{q}} \tag{20.8}$$

which measures the energy consumption due to the robot's motion relative to the total (robot + tooling) kinetic energy. A good tooling design would yield a large value of η, meaning that the effect of the tooling system is almost negligible, thereby addressing the issue of lightweight.

The second one, e_v, is the robot vibration ratio defined as

$$e_v = \frac{\omega_{0\mathrm{t}}}{\omega_0} \tag{20.9}$$

which evaluates the influence of the tooling system on the robot natural frequency, where $\omega_{0\mathrm{t}}$ and ω_0 are the fundamental natural frequency of the system with and without the tooling, respectively. It can be shown that $\sqrt{\eta}_{\min} \leq e_v \leq \sqrt{\eta}_{\max}$ [15]. Since η is less than 1, the tooling system will reduce the fundamental natural frequency, thereby addressing the issue of vibration.

The third one is the dynamic manipulability ellipsoid (DME) that measures the acceleration capability of the tool tip, thereby addressing the issue of compact size. The acceleration of TCP can be expressed as $a_t = J_t\ddot{q} + \dot{J}_t\dot{q}$. Assuming that the tool accelerates from rest, i.e., $\dot{q} = 0$, the TCP acceleration can be related to the joint actuation forces using Eq. (20.7) as [15]

$$a_t = B_t\tau + a_{gt} \tag{20.10}$$

where B_t is the matrix associated with the joint actuation forces, and a_{gt} is a vector associated with the gravitational acceleration. Therefore, the DME of the TCP can be evaluated by solving the singular values of matrix B_t. Three norms can be used: (1) $w_1 = \det(B_t)$, overall capability of TCP, where det(.) denotes matrix determinant; (2) $w_2 = \text{cond}(B_t)$, isotropy of DME, where cond(.) denotes the operation of matrix condition number; (3) $w_3 = \sigma_{min}$, indicating the lowest acceleration.

In terms of the aforementioned three indices, the two tooling designs as shown in Fig. 20.12 are compared. It can be seen from Table 20.2 that our tooling design for robotic percussive riveting yields a better performance than the traditional tooling design for automated squeezing riveting, because the index values of the former are overall higher than those of the latter.

In addition, a part feature-based method has been attempted to map sheet metal part features onto the tool approach direction (TAD). This mapping will not only help further tune the tooling design but also assist in determining correct directions for the tool to have a proper access for drilling and riveting. Generally, aircraft sheet metal parts can be classified in terms of bend direction and curvature. As shown in Table 20.3, sheet metal parts can be flat, single curved, and double curved
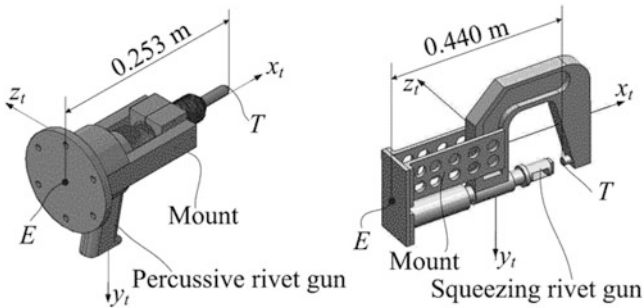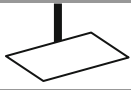
Fig. 20.12 Comparison of two tooling systems (percussive *left*, squeezing *right*)

Table 20.2 Comparison of two different tooling designs

Tool	η_{max}	η_{min}	e_v	w_1	w_2	w_3
Percussive	0.7265	0.9646	0.9436	71.3292	0.2161	1.7424
Squeezing	0.3489	0.8902	0.7928	27.4394	0.2940	1.4867

Table 20.3 Mapping of sheet metal shapes to tool's DOFs

Bend direction	Bend curvature (small)	Bend curvature (large)	Tool's DOFs
Flat		N/A	Three translations
Single curved			Three translations + one rotation
Double curved			Three translations + two rotations

Line Assembly — — — — — — — — — — — — — Airframe

Cockpit — Fuselage — — Wing — Empennage

Component Assembly

Sub-component Assembly — Forward fuse / Mid fuse / Aft fuse / Afterbody / Undercarriage — — Wingbox / Leading edge / Trailing edge / Wing tip / Engine mount — H Stabilizer / V Stabilizer / Elevator / Rudder
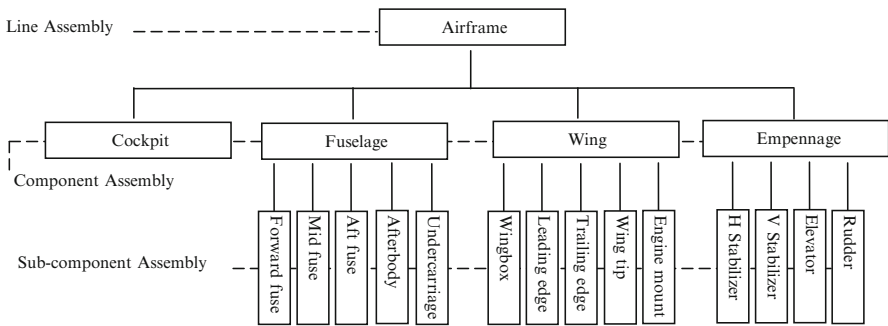
Fig. 20.13 Taxonomy of airframe assembly

shapes, each requiring 3, 4, and 5 degrees-of-freedom (DOFs), respectively, in order to position the tool normal to the surface. Though tool's DOFs are identical for both small and large curvatures, tool accessibility is different. The parts with large curvature would have tight spaces, difficult for the tool to access the rivet spot.

Tool accessibility is further affected by the areas surrounding the rivet spot. For this purpose, the taxonomy of a typical airframe assembly is provided in Fig. 20.13 to look at overall geometric constraints. It is true to say that fuselage and cockpit components involve parts with small curvatures, whereas wing and empennage components involve parts with large curvatures. Airframe is the main body of an aircraft, made of structural members and covered by skins. While bolts are used to join structural members, rivets are used to join skins. In other words, riveting process is mainly associated with aircraft skin assembly that may be divided into three steps: skin-stiffening, skin-joint, and skin-to-structure.

Skin-stiffening is to provide rigid support to a skin panel by riveting a number of stringers at the back of the panel. There are different shapes of stringer including Z, L, Y, I and hat-shape. Tool accessibility is affected not only by the size and geometry of the stringer but also the spacing between the stringers, as shown in Fig. 20.14a. Skin-joint is to join skin panels; there are two main joint designs,
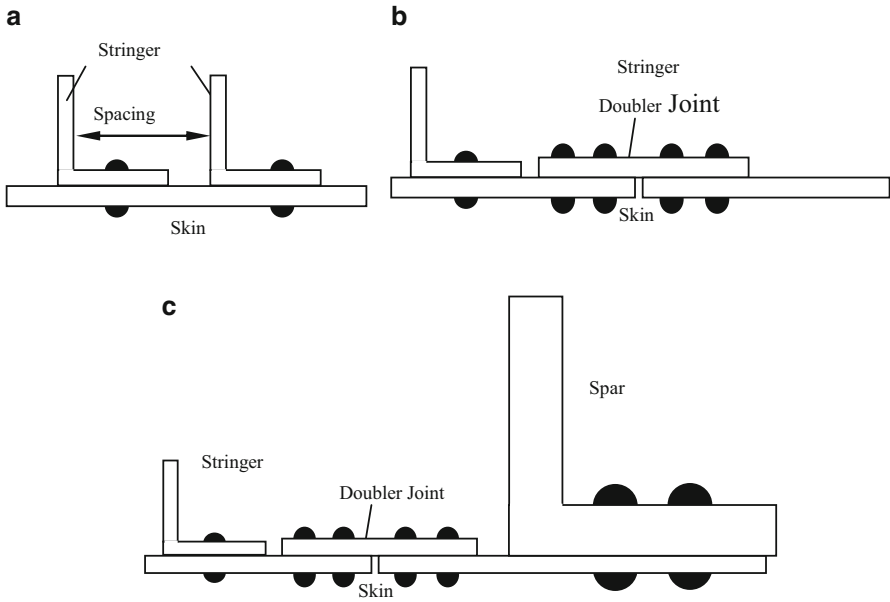
Fig. 20.14 Skin assembly. (**a**) Skin-stiffening. (**b**) Skin-joint. (**c**) Skin-to-structure

doubler (symmetric, as shown in Fig. 20.14b) and splice (asymmetric). Apparently, the size and geometry of the joint will also affect tool accessibility. Skin-to-structure is to mount skin panels onto a structure member, such as wing spar (as shown in Fig. 20.14c) or fuselage longerons (not shown). Probably, this is the most difficult part of skin assembly as the tool will be confined by the structures. In other words, as the skin assembly steps move up, tool accessibility becomes worsen.

3.4   Process Control

Process control is to study control methods for drilling and riveting. There are two main issues pertaining to robotic riveting, *localization* and *path tracking*. Localization is to transfer the coordinates of the rivet spots to those in the robot frame. As shown in Fig. 20.15, a position sensor system is used to measure both the jig and the rivet gun. By measuring three tooling balls attached to the jig, the jig frame, denoted by $^s\mathbf{H}_j$, can be determined using a three-point method [16]. Likewise, by measuring three makers attached to the tool, the tool frame, denoted by $^s\mathbf{H}_t$, can also be determined. Then the jig frame can be expressed with respect to the tool frame as

$$^t\mathbf{H}_j = (^s\mathbf{H}_t)^{-1}\,{}^s\mathbf{H}_j \tag{20.11}$$
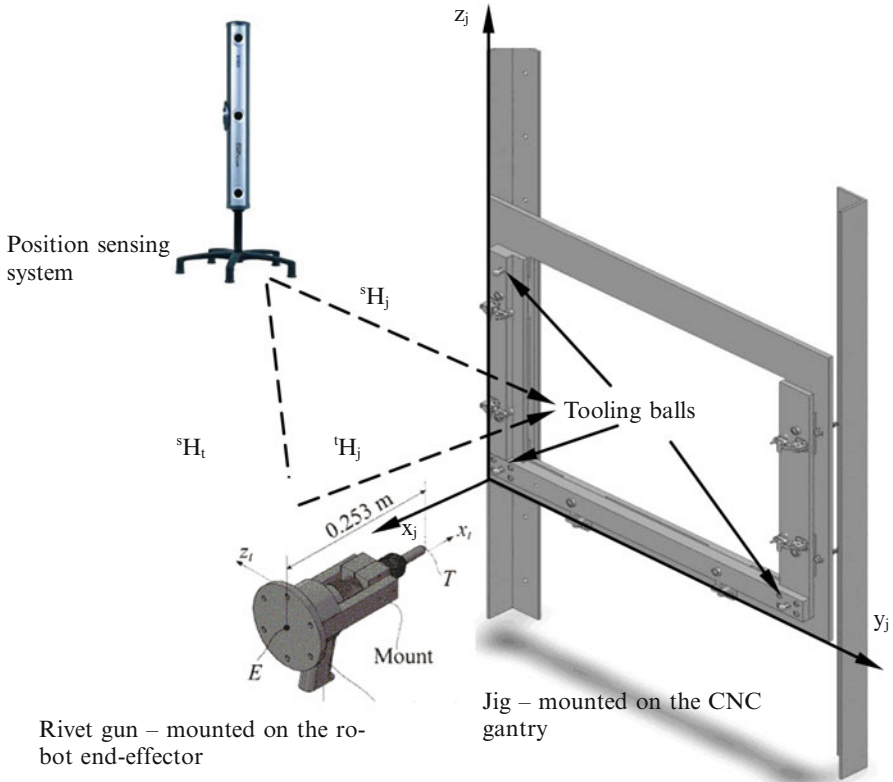
Fig. 20.15 Localization

where $^x\mathbf{H}_y$ represents a homogeneous transformation matrix from a y frame to a x frame, subscripts s, t, and j stand for sensor frame, tool frame, and jig frame, respectively. Since the rivet spots are expressed with respect to the jig frame, they can be readily transferred to the tool frame using Eq. (20.11), based on which the robot can be programmed to follow these spots.

Path tracking is to control the path of tool tip to reach each rivet spot precisely based on visual sensing. As mentioned before, the riveting process adopted here is sequential, i.e., drilling all the holes first based on a planned path, and then riveting them along the same path. Though the identical path program is used for both, there are two main sources of error needing advanced control for riveting. One is robot path positioning accuracy and repeatability, and another is hole tolerance. Despite calibration to reduce systematic errors, industrial robots still exhibit a repeatability problem due to various uncertainties (random errors). Temperature changes could cause robot drift. Further, the robot calibration is normally done within entire workspace by robot manufacturers. In the local working zone, the path positioning accuracy may exceed the overall accuracy and deteriorate the rivet insertion. For

the robot used in our riveting system, it has a path repeatability of 0.6 mm and a position repeatability of 0.2 mm. In general, the hole is drilled with a tolerance of 0.1–0.2 mm bigger than the rivet in diameter. Apparently, the robot repeatability would not be able to guarantee each time the successful insertion of a rivet from the tip of the rivet gun to inside the hole. Unsuccessful rivet insertion would cause damage to sheet metal skins or to the tooling system. For this reason, advanced controls are investigated.

Our control method consists of two parts, one dealing with the rivet insertion path and another dealing with the hole geometry. The first part is to carry out a continuous relative pose estimation of ${}^{t}\mathbf{H}_{j}$ based on Iterative Extended Kalman filter (IEKF) method [17] and then use it as estimation error for path tracking based on Iterative Learning Control [18]; the goal is to keep track of the tool pose in the course of insertion. On the other hand, we cannot change the kinematic parameters in all industrial robot controllers due to its not-open architecture; hence, the solution of process control is to develop a path corrector that can correct the preplanned path through ILC for the improvement of path tracking accuracy for rivet insertion.

The problem can be described as discrete time-varying nonlinear stochastic systems in state space domain,

$$X(k+1) = FX(k) + G(k)u(k) + \Gamma v \tag{20.12}$$

$$Y(k) = h(X(k)) + Mw \tag{20.13}$$

where $X(k)$ is the state variables, including three variables for the tool position and other three for the tool orientation, and F is a state transition matrix. $u(k)$ is a control input with a control gain of $G(k)$, $Y(k)$ is a measurement vector provided by the aforementioned position sensing system, and $h(X(k))$ is a nonlinear observation mapping model from the state to the measurement. Γ and M are the gains for noises of v and w, respectively.

Based on the model, the IEKF is applied to estimate the state mean and covariance of tool pose along the rivet insertion path,

$$\widehat{X}^{i}(k,k) = \widehat{X}(k,k-1) + K^{i-1}(k)\left[e^{i-1}(k) - H^{i-1}(k)\left(\widehat{X}(k,k-1) - \widehat{X}^{i-1}(k,k)\right)\right] \tag{20.14}$$

$$P(k,k) = \left(I - K^{i}(k)H^{i}(k)\right)P(k,k-1), \tag{20.15}$$

where $\widehat{X}^{i}(k,k)$ and $P(k,k-1)$ are the mean and covariance the estimated tool pose, respectively, $K^{i-1}(k)$ is the Kalman gain, $e^{i}(k)$ is the measurement residual, and $H^{i}(k)$ is the linear approximation of $h(X(k))$. Figure 20.16 shows a simulation result of the relative pose estimation to demonstrate the effectiveness of the method.
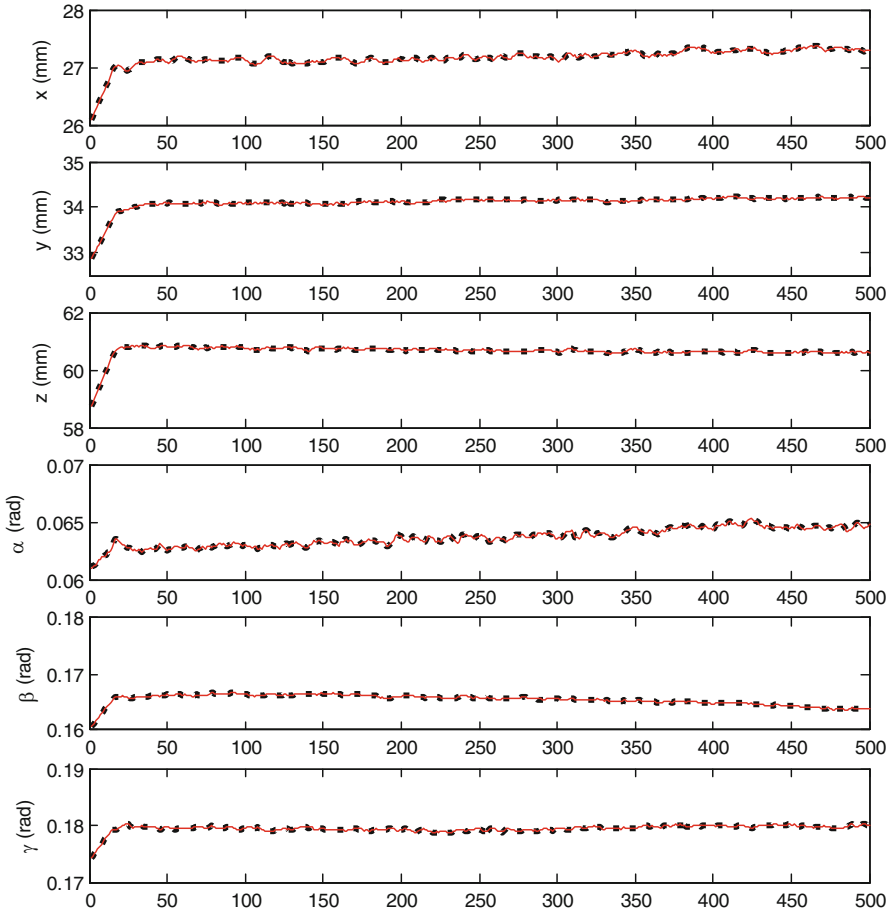
Fig. 20.16 Kalman filter method—*red solid curves* represent the nominal values of relative pose, whereas *black dash plots* represent the estimated values by Kalman filter

For path tracking, the differences between desired and estimated tool pose are calculated as e_i and then applied in the updating law of ILC for a robot local working zone-calibrated kinematic parameters as

$$p_{i+1} = p_i + C_i \exp^{(-rt)} e_i \qquad (20.16)$$

where C_i is the learning gain matrix of the path tracking, r is a positive constant, and t is the learning time in the ith iteration. The convergence property of p_i approaching p is proved to ensure the path error is reduced.

After the kinematic parameter identification through the ILC scheme along the path in a working zone, the actual (local) kinematic parameters become p. The kinematic parameter difference is given by

Robotic percussive
riveting cell
- Controller

CNC bucking bar gantry - Controller

Tooling system - Controller
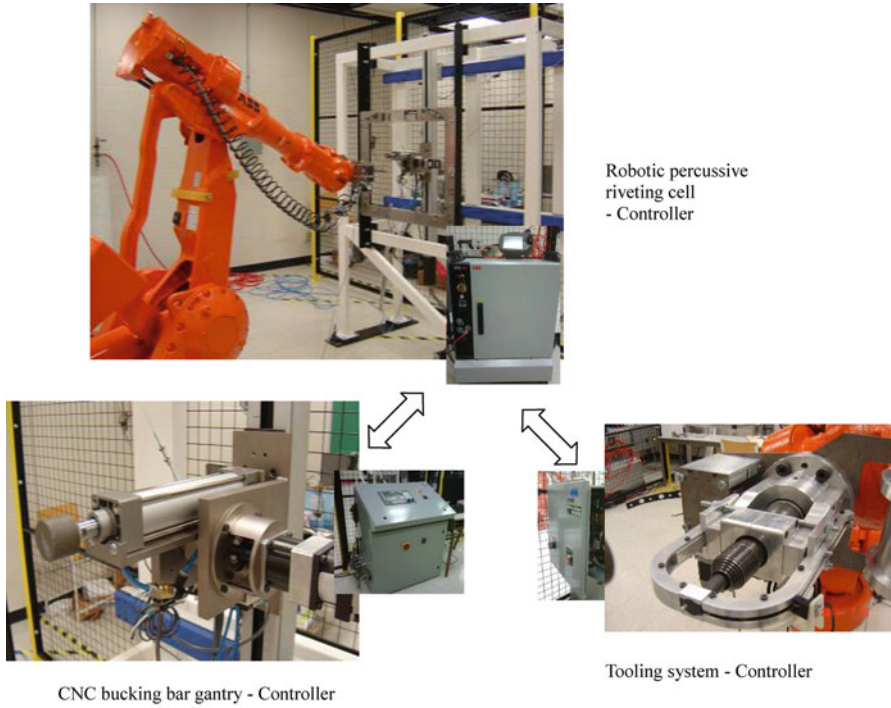
Fig. 20.17 Robotic percussive riveting system
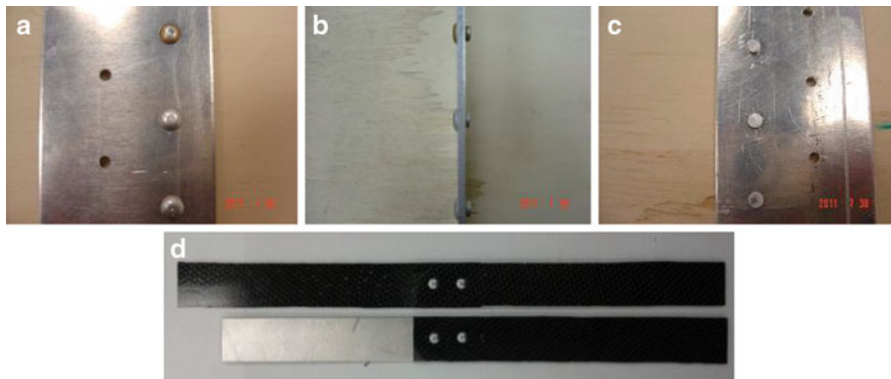
Fig. 20.18 Robot riveted samples. (**a**) front—metal, (**b**) side—metal, (**c**) back—metal, (**d**) composite

$$\delta p = p - p_r \tag{20.17}$$

Then, the values for path correction via the following relation:

$$\delta y = F_p(x_r, p)\,\delta p \tag{20.18}$$

where F_p represents the commonly called geometric error mapping matrix. At the end, $y = y_r + \delta y$ will be actual input to the robot controller, leading the robot to the required pose of rivet insertion.

The second part is to determine accurately the position of the hole center in the jig frame, because though the holes are drilled according to a planned path, they will deviate due to various errors. This is a onetime measurement using a high-resolution camera mounted on the robot. Efforts have been devoted to improve the accuracy of the hole center position that is computed from a digital image. This work is also being applied to the tooling system self-calibration. Figures 20.17 and 20.18 show the developed robotic riveting system and test samples.

3.5 Robot Interaction

An important issue pertinent to robot interaction during percussive riveting is the robot fatigue resulting from the repetitive impacts. A similar problem has been addressed for manual operations [19]. During riveting, repetitive percussions act as vibration excitations to the robot. For industrial robots, their links are made relatively bulky, and hence can be considered as rigid. However, their joints are relatively weak, and usually considered flexible. Percussive riveting operation will cause robot joint vibrations, hence leading to joint fatigue problem.

For flexible-joint robots, modeling is done using linearly elastic torsional springs. If we choose the input joint angles as the generalized coordinates, i.e., $\mathbf{q} = [\theta_1, \theta_2, \ldots, \theta_n]^T$, the kinetic and elastic energies of a stationary flexible-joint robot can be respectively written in terms of the mass and stiffness matrices as

$$T = \frac{1}{2}\Delta\dot{\mathbf{q}}^T\mathbf{M}(\mathbf{q})\,\Delta\dot{\mathbf{q}}, \quad K = \frac{1}{2}\Delta\mathbf{q}^T K \Delta\mathbf{q} \tag{20.19}$$

where $\Delta\mathbf{q} \in \mathbf{R}^n$ is the perturbations in joint angles from an equilibrium configuration $\mathbf{q}_0$, representing the joint deflections, $\Delta\dot{\mathbf{q}}$ is the deflection rate, $\mathbf{M}(\mathbf{q})$ is the $n \times n$ symmetric generalized mass matrix, and $\mathbf{K}$ is the $n \times n$ diagonal stiffness matrix in the joint space, with its entries equal to the torsional stiffness of the corresponding joints. Then, the dynamic model when the robot is the stationary and under forced vibration can be represented as [20]

$$M(q)\Delta\ddot{q} + K\Delta q = F(t)J^T n \tag{20.20}$$

where the $3 \times n$ Jacobian matrix $\mathbf{J}$ represents the mapping from the joint linear velocities to the velocities of tool center, and $\mathbf{n}$ is the direction vector of the riveting force.

Once the joint deflections are obtained by solving Eq. (20.20), the shear strain of the joint shaft can be calculated as

$$\gamma = \frac{\Delta q r}{l} \tag{20.21}$$

where r is the radius of the joint shaft and l is the shaft length. Now the shear stress of the joint shaft can be obtained as follows:

$$\tau = G\gamma \tag{20.22}$$

where G is the shear modulus of the joint shaft.

The fatigue life of the robot joints under torsional vibrations can be estimated from the shear stresses of the joint shafts. In practice, the robot will perform riveting at different positions in different directions over the workspace. It is well known that the robot joint forces are configuration dependent. In other words, even under the same percussive riveting forces, the joint vibrations will differ and the joint shear stresses will vary. Therefore, a multiple-loading fatigue model must be applied. As shown in Fig. 20.19, different shear stresses will produce different number of cycles to failure. To account for this, the Palmgren–Miner cycle-ratio summation rule, also called Miner's rule, is adopted here.

$$\sum_{i=1}^{k} \frac{n_i}{N_i} = c \tag{20.23}$$
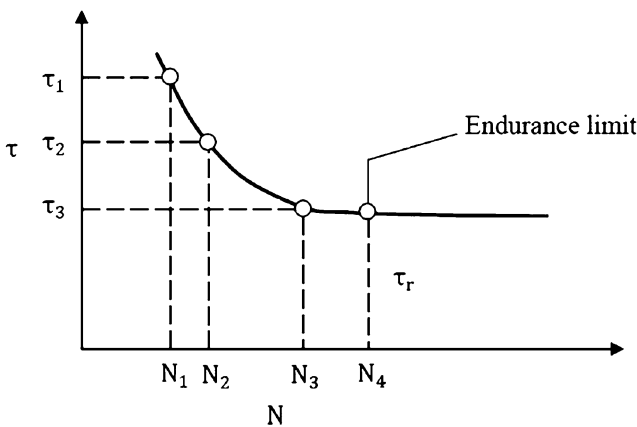
Fig. 20.19 Multiple-loading fatigue S–N curve
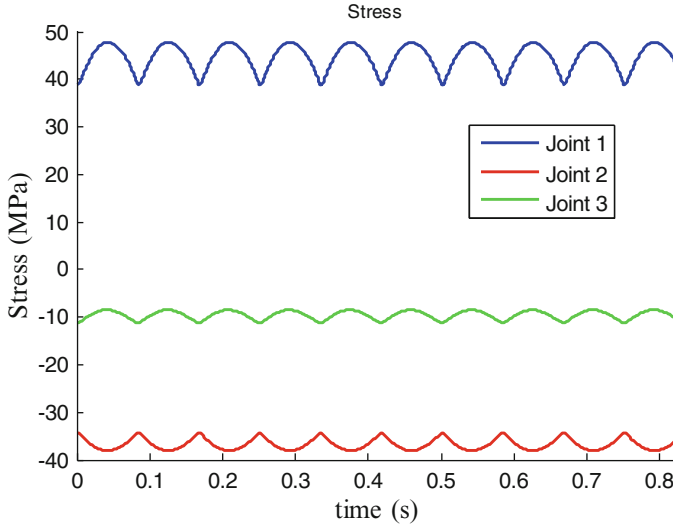
Fig. 20.20 Stress of the flexible joints during percussive riveting

where n_i is the number of cycles at stress level τ_i, N_i is the number of cycles to reach failure at stress level τ_i, and k is the number of different riveting operations. The parameter c is determined by experiment and found in a range of between 0.7 and 2.2. Usually, $c = 1$ is used.

Simulation is carried out to investigate the joint stresses and fatigue life under percussive riveting. Figure 20.20 shows a typical joint shear stress under percussive riveting. It can be seen that the joint stress fluctuates. To determine the fatigue life, an S–N curve of the used material is needed. For this study, the S–N curve is obtained through fitting the data for steel. This S–N data was generated from the experiment for the material similar to those used for common robot joint shafts. The equation for the fatigue curve is

$$\tau = aN^b \tag{20.24}$$

By fitting it is found that $a = 18930$ and $b = 0.3548$.

In our simulation using Matlab, a riveting task including three directions, forward, left, and upward, is considered for fatigue life prediction. The same rivet pattern as shown in Fig. 20.21 is applied to all three directions. There are in total 15 rivets and the numbers in Fig. 20.21 indicate the sequence of riveting, i.e., the rivet path. For all riveting the same size of rivets is used. To predict robot fatigue life, the fatigue stress of each direction must be determined to find the corresponding cycle numbers to failure. The number of hits per rivet is found through experiment to be 12. Figure 20.22a shows the riveting in the forward direction. The impact force direction n in this case is $\mathbf{n} = [1, 0, 0]^T$. Figure 20.22b shows the maximum stresses of the three joints overriveting 15 rivets at 15 rivet spots. The mean value of the
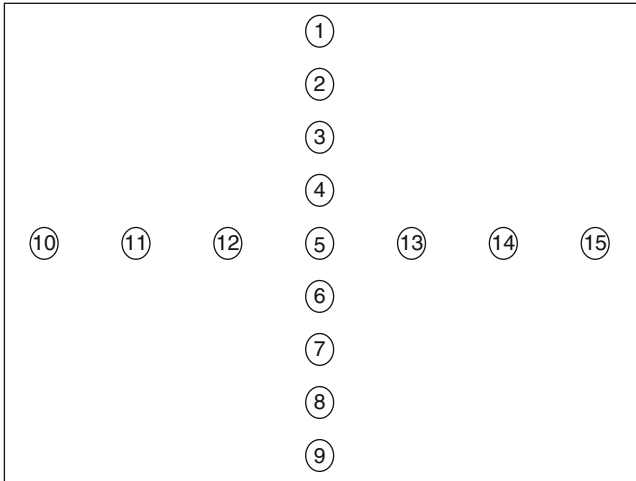
Fig. 20.21 Rivet pattern and corresponding path
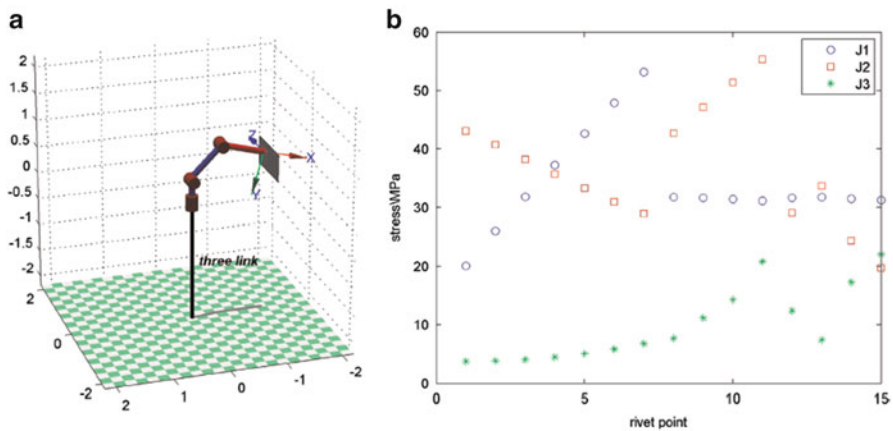
Fig. 20.22 (**a**) Riveting forward and (**b**) corresponding joint stresses

fatigue stresses over 15 rivets are found as 32.81, 37.46, and 9.4 3 MPa, which are used to find N_i for the three joints, respectively. Figure 20.23a depicts the riveting in the left direction, with the impact force direction $\mathbf{n} = [0, 1, 0]^T$. Figure 20.23b shows the maximum stresses of the three joints over 15 rivets. The mean value of the fatigue stresses over 15 rivets are found as 35.56, 37.91, and 11.13 MPa for the three joints, respectively. Figure 20.24a displays the riveting in the left upward direction, and the impact force direction $\mathbf{n} = [0, 0, 1]^T$. Figure 20.24b shows the maximum stresses of the three joints over 15 rivets. The mean value of the fatigue stresses over 15 rivets are found as 0, 54.86, and 56.23 MPa for the three joints, respectively. Now, the total number of rivets that can be riveted over the robot's fatigue life can
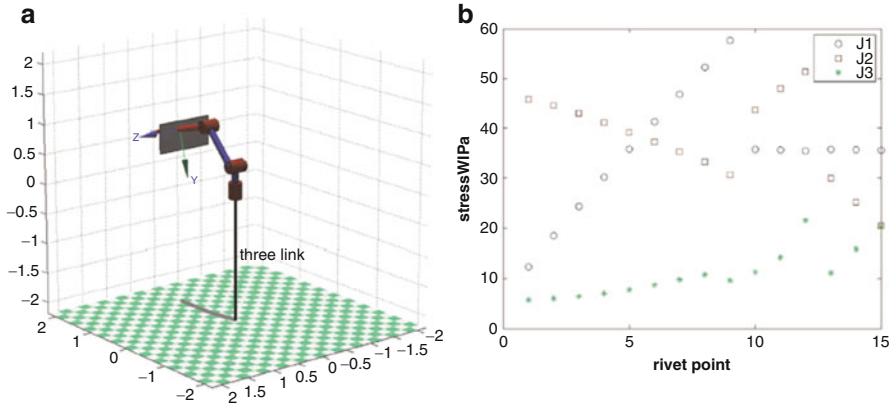
Fig. 20.23 (**a**) Riveting left and (**b**) corresponding joint stresses
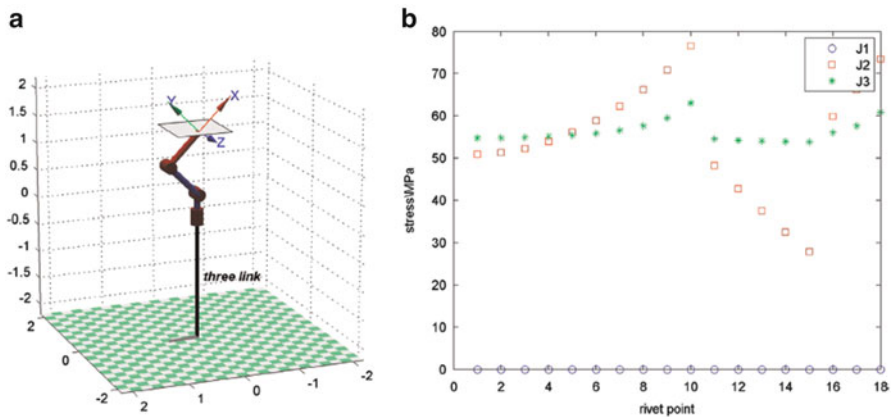
Fig. 20.24 (**a**) Riveting upward and (**b**) corresponding joint stresses

be determined from Eq. (20.23). For joint 1, n is 355,384; for joint 2 and 3, n are 144,797 and 253,420, respectively. Finally, the smallest value among the three joints is used to predict the robot's fatigue life, i.e., there are 144,797 rivets that can be riveted before the robot's fatigue failure occurs.

4 System Implementation and Concluding Remarks

The research results presented in this chapter have been applied to develop a robotic percussive riveting system. Figure 20.18 shows the physical system involving three controllers for three subsystems, one for the robot, one for the tooling, and one for the bucking bar gantry. All three controllers are integrated, with the robot

controller being the main controller for synchronization. A complete riveting control sequence has been generated, starting from position the gun → position the bucking bar → insert rivet → extend the bucking bar → rivet → retract the gun → retract the bucking bar → move to the next spot. This control sequence has been successfully tested and implemented to perform percussive riveting on sheet metal panels and composite panels as shown in Figure 20.18.

Our experience gained through this development has clearly indicated that a successful robot application to the automation of a process requires in-depth research on the process and the interaction with the robot. The said research can be systematically carried out according to key process planning considerations including: process sequence, process parameters, process tooling, and process control. It has been demonstrated that through this list, a number of key research issues were identified for robotic percussive riveting including rivet pattern planning, rivet time determination, rivet tooling design, and rivet insertion control. Furthermore, an important issue pertinent to robot interaction is identified, i.e., robot fatigue life under repetitive percussion during riveting. The detailed research on these issues has led to know-how for the successful implementation of our robotic percussive riveting system.

References

1. Campbell FC (2006) Manufacturing technology for aerospace structural materials. Elsevier, Amsterdam, pp 495–537
2. Mortimer J (2001) Jaguar uses X350 car to pioneer use of self-piercing rivets. Ind Robot 28(3):192–198
3. Morey B (2007) Robot seeks its role in aerospace. Manuf Eng 139(4):AAC1–AAC6
4. Inman J, Carbrey B, Calawa R, Hartmann J, Hempstead B, Assadi M (1996) Flexible development system for automated aircraft assembly. In: SAE aerospace automated fastening conference & exposition, Bellevue, WA, Oct 1996
5. Kleebaur R (2006) Where precision counts above all else. High Flyer 2:12–14
6. Monsarrat B, Lavoie E, Cote G, De Montigny M, Corbeil C, Grenier D, Tu X, Perron C (2007) High performance robotized assembly system for challenger 300 business jet nose fuse panels. In: AeroTech 2007, Los Angeles, CA
7. Kim SJ, Paik SH, Ji KH, Yoon TH (2007) 3D riveting process simulation of laminated composites. Key Eng Mater 334–335:405–408
8. Groover MP (2004) Fundamentals of modern manufacturing, materials, processes and systems, 2nd edn. Wiley, New York
9. Skorupa M, Korbel A (2008) Modeling of the secondary bending in riveted joints with eccentricities. Arch Mech Eng IV(4):369–387
10. AC 43.13-1B—acceptable methods, techniques, and practices—aircraft inspection and repair. Federal Aviation Administration of USA. Issued on 8 Sept 1998
11. Bhandari LB (2006) Introduction to machine design. Tata McGraw-Hill, New Delhi

12. Yoon TH, Kim SJ (2011) Refined numerical simulation of three-dimensional riveting in laminated composites. J Aircr 48(4):1434–1443
13. Li Y, Xi F, Behdinan K (2010) Modeling and simulation of percussive riveting for robotic automation. ASME J Comput Nonlinear Dyn 5(2):021011
14. Jenkins H (2005) Design of robotic end effectors. In: Kurfess TR (ed) Robotics and automation handbook. CRC, Boca Raton
15. Li Y, Xi F, Mohamed R, Behdinan K (2011) Dynamic analysis for robotic integration of tooling systems. ASME J Dyn Syst Meas Control 133(4):041002
16. Lin Y, Tu X, Xi F, Chan V (2013) Robust pose estimation with an outlier diagnosis based on a relaxation of rigid body constraints. ASME J Dyn Syst Meas Control 135(1):014502
17. Lin Y, Chen T, Xi F, Fu G (2015) Relative pose estimation from points by Kalman Filters. In: 2015 IEEE conference on robotics and biomimetics
18. Zhao Y, Lin Y, Xi F, Guo S (2015) Calibration-based iterative learning control for path tracking of industrial robots. IEEE Trans Ind Electron 62(5):2921–2929
19. Cherng JG, Eksioglu M, Kizilaslan K (2009) Vibration reduction of pneumatic percussive rivet tools: mechanical and ergonomic re-design approaches. Appl Ergon 40(2):256–266
20. Nie X, Li Y, Guo S, Song T, Xi F (2016) Modeling and simulation for fatigue life analysis of robots with flexible joints under percussive impact forces. Robot Comput Integr Manuf 37:292–301

Chapter 21
Photo-Induced Fabrication Technology for 3D Microdevices

D. Serien, Y. Morimoto, and S. Takeuchi

Abstract This chapter introduces photo-induced fabrication technologies for 3D MEMS devices. At first, we introduce principles of 3D photo-induced fabrication focusing on common aspects of photo-induced cross-linking and range of applicable materials. Then, we examine four technologies and their outcome of applications in detail where fabricated feature sizes decrease and resolution increases with progression of this chapter. (1) Microstereolithography enables a layer-by-layer fabrication of 3D devices, which find application in coaxial microfluidics and device fabrication in the presence of cells. (2) *In situ* photolithography is a 3D device fabrication with a pre-polymer solution inside microfluidic devices enabling fabrication of 3D structures inside microfluidic channels. (3) Flow lithography applies the concept of flowing pre-polymer material inside a microfluidic device in order to fabricate variously shaped microparticles for self-assembly or potential use as drug delivery systems. (4) Direct laser writing exposes focused laser light moved by computerized piezo actuation through pre-polymer medium enabling to structure with nanometer-sized feature sizes and constructing materials with special optical or mechanical properties as well as 3D devices for cell culture studies. The application of the achieved 3D devices ranges from microfluidic elements over bioMEMS devices constituting of analytical tools, cell culture conservation, and cell culture studies to biomedical application suitable for implants.

Keywords Microstereolithography • *In situ* photolithography • Flow lithography • Direct laser writing • Multiphoton • Microparticle • Cell culture scaffold • Metamaterial • Photoreactive • Photo-initiated

D. Serien • Y. Morimoto • S. Takeuchi (✉)
Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Takeuchi Biohybrid Innovation Project, Exploratory Research for Advanced Technology, Japan Science and Technology, Tokyo, Japan
e-mail: takeuchi@iis.u-tokyo.ac.jp

1   Introduction

Photo-induced technologies for patterning materials are a fundamental concept to fabricate MEMS devices [1, 2]; Photo-induced polymerization is a technology based on light-activated polymerization in photoreactive material by various light sources. Conventional photolithography utilizes light to polymerize exposed so-called negative photoresists or to make exposed, the so-called positive photoresists soluble for consecutive developing steps. It can be a goal to shape the photoresist for example to use it as mold for microfluidic device casting in PDMS [3], but also using the photoresist as a coating or etching mask for underlying materials or sacrificial material [4, 5]. By these means, various materials can be structured including complicated features by gray-scale [6, 7] and hollow elements by interference [8, 9], moving mask [10] and defocusing [11]. However, the layer-to-layer concept increases difficulty with each layer due to the necessity of alignment and proper layer adhesion. One challenging aspect of layer-to-layer fabrication is consideration of processing step specificity to prevent undesired interaction of individually targeting processes with all device materials. For instance, shaping multiple layers requires multiple mask fabrication steps and etching technologies that mainly access the target layer without considerably altering already fabricated structure elements. Then, fabrication of a complex channel shape or freely moving entrapped elements is an intensive challenge.

The fabrication of 3D elements or entire 3D MEMS devices that are not bound to the layer-to-layer planning of conventional lithography is important to enable optically driven microfluidic pumps [12], 3D coaxial microfluidics [13, 14] as well as fabrication of elements and modules for further assembly to bottom-up devices [15, 16]. In order to fabricate such 3D elements, other fabrication technologies emerged from conventional photolithography utilizing the photoreactive concept.

The goals of 3D microdevice fabrication are appealing to multiple fields. In microfluidics, a circled cross section results in ring-shaped flow-focused position of particles in contrast to a squared cross section resulting in a wall-face-focused position, and similarly due to inertial flow, channel curves, channel joining angle and channel asymmetry can be used for diffusive mixing, flow-focusing, and particle filtering [17]. Additionally, the integration of freely moving elements inside the microfluidic device enables fabrication of gates [18] and pumps [12, 19] via the introduction of photoreactive material and performing exposure processes inside microfluidic devices, the so-called *in situ* photolithography. Furthermore, 3D microdevice fabrication is purposed to complement microfluidic technology with 3D device elements that enable round channels, channel into channel integration, and conventional fluid elements such as glass capillaries, tubes, syringes, and pumps [13, 14]. Moreover, 3D microfabrication is the applied towards cell culture studies and tissue engineering by fabricating structures for cells to interact or grow on (cell culture scaffold) [20], by fabricating structures laden with cells that are portable for further arrangement [5, 14, 21], and by fabricating structures around target cells to guide or capture them for further study [22, 23]. In addition, 3D

microdevice fabrication is extensively utilized for the fabrication of optical [24], and mechanical [25, 26] metamaterials, materials that because of their internal structure exhibit exceptional properties [27, 28]; Photoreactive material is structured down to submicron feature sizes controlling the shape of individual elements to wire, beam or asymmetric diamond-like shaped elements and controlling the connectivity of these material elements to how many elements meet in which angle per connectivity nod in 3D to create mesoscale networks.

In this chapter, we will (1) introduce fundamental principles underlying photo-induced fabrication of 3D microdevices and (2) review state-of-the-art fabrication technologies for 3D microdevices with examples of device application.

2 Principles of Photo-Induced Fabrication of 3D Microdevices

In this section, we will briefly cover principles and ideas that are common amongst the methods introduced in later sections and required to understand the basics of photo-induced fabrication.

A general fabrication concept is schematically illustrated in Fig. 21.1. Light is exposed from a light source. With computer-assisted control, light is patterned by
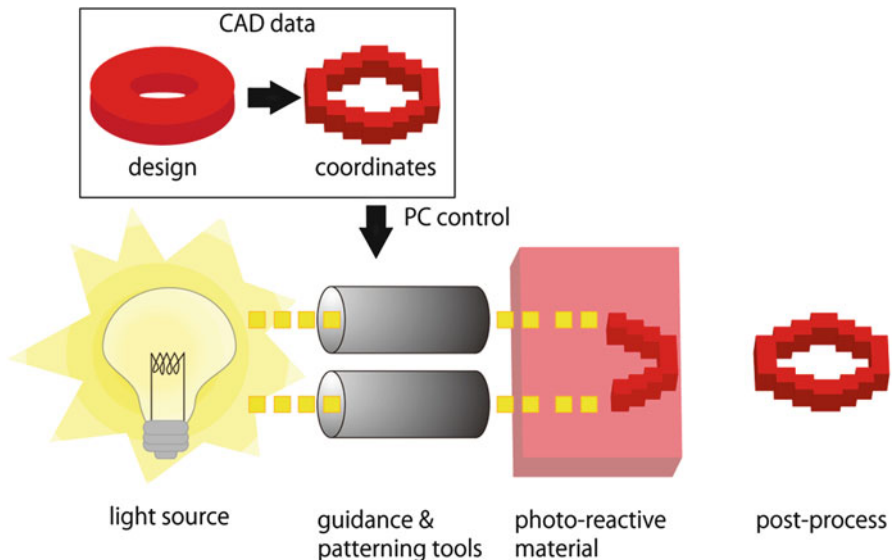
Fig. 21.1 Abstract workflow principle of photo-induced fabrication for 3D microdevices. Light from a light source is guided or masked in accordance with PC control regarding the coordinates derived for a structure design. Light interacts with photoreactive material. After appropriate processes such as development, rinsing, or functionalization, the goal structure is obtained

guiding, masking or modulating tools. The computer-assisted manipulation of light requires knowledge about the manipulation and polymerization processes in order to obtain the desired structure. When light interacts with photoreactive material, areas with polymerization and areas without polymerization are created. After further chemical treatment (development) and heat treatment (baking), washing (rinsing), or optional surface alteration for special purposes (functionalization), the 3D structure is obtained.

2.1 Light Exposure

Regardless the light source, commonly Halogen lamp, LED or laser, photo-induced fabrication for 3D microdevices requires light excitation, i.e., changing the energy state of a target molecule by light absorption to an alleviated state (Fig. 21.2) [1, 29]; When the energy transferred by light as photons matches the energetic difference of energy levels, molecules absorb the energy alleviating their energy state, commonly from ground state to an energetically higher excited state (Fig. 21.2a). Commonly, the molecule in excited state contributes to the start of the polymerization initiation and relaxes with or without recovery after a certain lifetime in an intermediate state [1, 29]. Alternatively, the molecule may decompose due to its excitation and yield the actual active component contributing to polymerization as product of the decomposition [30].
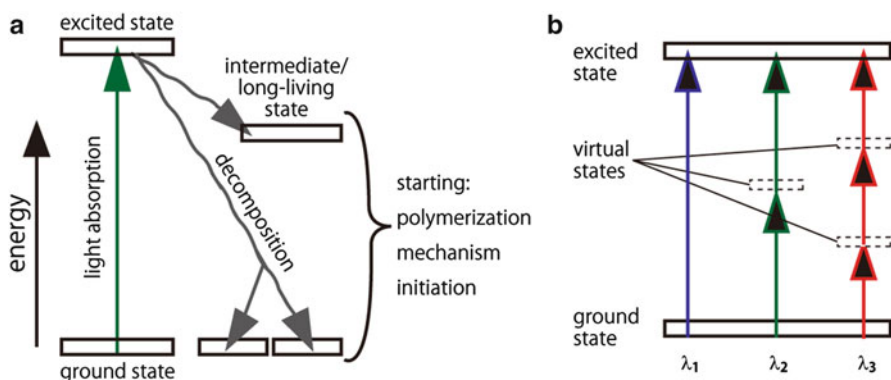
Fig. 21.2 Diagram for illustration of light excitation. (**a**) After light is absorbed, the absorbing molecule enters an energetically alleviated excited state. The polymerization mechanism initiation is started by molecules that relax from the excited state to an intermediate/long-living state or by molecules that are generated by decomposition of the absorbing molecule. (**b**) Comparable energetic excitation is achieved by one-photon excitation of wavelength λ_1, two-photon excitation of wavelength $\lambda_2 = 2\lambda_1$, or three-photon excitation of wavelength $\lambda_3 = 3\lambda_1$. For multiphoton excitation, photons are absorbed simultaneously, expressed by the inhabitation of virtual energetic states

The required energy for excitation is dependent on the pre-polymer material absorption spectrum, but the choice of experimental wavelength is dependent on whether the fabrication is aiming to exploit one-photon, two-photon, or three-photon excitation (Fig. 21.2b); coincidental absorption of multiple photons enables to cover the comparable energetic excitation of one-photon excitation. One-photon absorption is linearly dependent on light intensity, leading to quick polymerization of large areas or volumes compared to multiphoton excitation. Two-photon absorption is quadratically dependent on light intensity leading to a slimmed exposure profile, localization of exposure light to a focal volume and improved polymerization resolution [1, 31]. Photons are absorbed simultaneously, expressed by the inhabitation of virtual energetic states. Any combination of energetic values that achieve comparable energetic sum could potentially be used for multiphoton processes [31, 32]. Multiphoton excitation wavelength are usually chosen to be degenerated, i.e., a multiple of the energetically identical wavelengths fulfill the excitation energy requirement [18, 31, 32], for the practical reasons of using one wavelength light source and same wavelength diffraction leading to simple foci alignment. In order to achieve higher occurrences of coincidental photon absorption, fs-pulsed lasers are commonly used because they exhibit high photon density during pulses in contrast to continuous-wave lasers [1].

2.2 Light Patterning Tools

In order to create complex 3D features by devising polymerized and non-polymerized volumes, light patterning is required. Light patterning can be classified into three distinct approaches: hard mask, dynamic mask, and direct writing (Fig. 21.3).
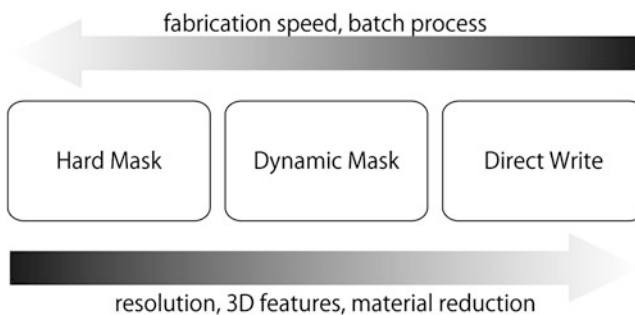
Fig. 21.3 Comparison of light patterning tools. Choice of light patterning tools towards hard masks and conventional photolithography enables batch processes and increased fabrication speed, while direct light exposure enables light-dependent resolution and 3D feature control. Dynamic masks for phase or amplitude manipulation trade-off by dynamically generating patterns layer-to-layer

Usage of hard masks, i.e., a metal or glass device shaped to provide the desired patterning, is originated from conventional photolithography technology [1]. Once fabricated, a hard mask can be used repeatedly without additional processing time. Well-designed masks can be used for batch fabrication. The initial fabrication cost is comparably high. For fabrication with truly 3D structure features i.e., arbitrary design in lateral as well as in axial direction, fabrication cost becomes unreasonable and processing time increases with mechanical mask exchange [33].

Dynamic masks are electrically actuated devices where pixels can be accessed individually in order to generate patterns dynamically. Dynamic masks therefore cancel out fabrication costs of hard masks, while harnessing a quick processing time where not the device but its pixel orientation changes for pattern generation. Lateral and axial shift can be further modified to match image plane, enabling even quicker 3D patterning. Occasionally, these technologies are considered maskless technology because they do not require fabrication of physical masks [33, 34]. Dynamic-mask technology commonly utilize either dynamic micromirror devices (DMDs), where each pixel is an individually actuated micrometer-sized mirrors, (Fig. 21.4a, c) [23, 34–38] or liquid crystal displays (LCDs), where actuated crystal orientation per pixel area determines light transmission via polarization, (Fig. 21.4b, d) [7, 39]. There are two different concepts in use to generate an effective light intensity
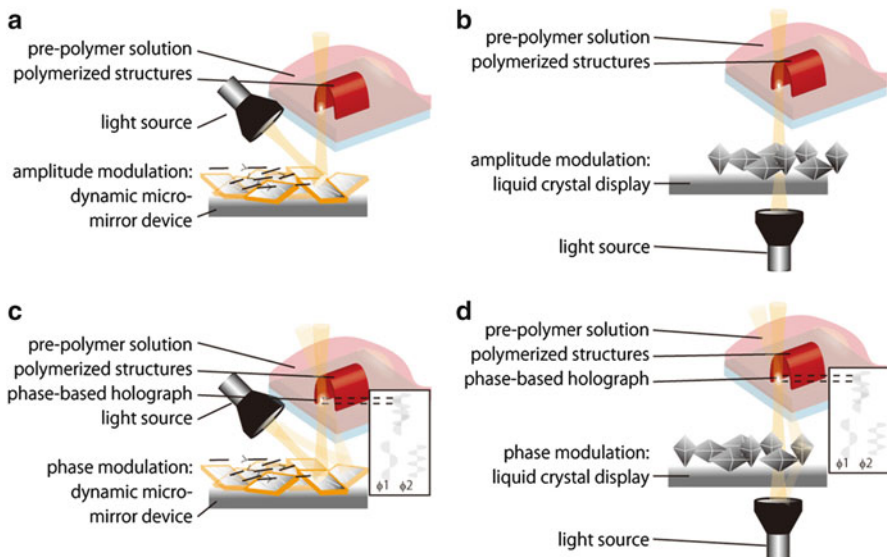
Fig. 21.4 Dynamic-mask lithography. In the light path, a device manipulates passing light representing a programmed pattern. (**a**, **b**) Optical dynamic-mask lithography: the device, either (**a**) a dynamic micromirror device (DMD) or (**b**) a liquid crystal display (LCD), is used to modulate the amplitude of light. (**c**, **d**) Holographic dynamic-mask lithography: the device, either (**c**) a DMD or (**d**) a LCD, is used to modulate the phase of light. Adjusting focus position, layer-stacking is achieved

pattern per layer, amplitude-based optical pattern generation (Fig. 21.4a, b) [7, 23, 35, 36, 39] and phase-based holographic pattern generation (Fig. 21.4c, d) [34, 37, 38]. For optical projection, the digital mask generating device is interrogated by a raster-scanning laser. Multifocal fabrication is achieved by first directing the laser beam through a beam splitter [40] or utilization of an optical fiber array [41]. For holographic projection, phase modulation required for the first diffraction pattern is determined by an algorithm, such as direct binary search or the Gerchberg–Saxton algorithm [37]. The diffraction pattern enables parallel fabrication comparable to multifocal lithography where complexity and number of focal points are in trade-off with diffraction pattern intensity, influence of phase irregularities and noise [34]. Based on the device pixel actuation and the layering concept, dynamic masks are limited in resolution size and in fabrication type of layer-to-layer or one-shot.

Direct laser exposure to the pre-polymer solution enables polymerization down to the diffraction limit of the corresponding excitation. Multiphoton excitation especially enables submicron feature sizes [1, 29]. As maskless exposure, the laser is moved by a piezo stage accessing all Cartesian coordinates, leading to the possibility to fabricate along an arbitrary path. Thereby, raster-scanning, movement along a fixed raster with shutter-timed exposure control, and vector-scanning, movement along a computed vector combined with exposure control, both become available [42], as well as a combination of these two scanning modes are feasible. General resolution criteria can be even surpassed when utilizing the so-called super-resolution; similar to fluorescence microscopy stimulated emission depletion (STED), reduction of spot size resolution is achieved by depletion of the environment of an excited spot by a second light-emission-inducing exposure [29].

2.3 Materials

The variety of photoreactive materials is large, ranging from materials derived from conventional photolithography such as SU-8 photoresist [34] and PDMS [30], commercial products such as resins [13] and optical adhesives [43] to specially synthesized chemicals such as thiol-ene-based polymers [44] (Table 21.1, Material compilation).

Despite the variety, common properties can be identified. Trivially, the material is requires to be photoreactive and transmitting at least the wavelength required for polymerization. Most materials also exhibit viscosity, but while it is desired in some cases [41], by adding diluents viscosity can be lowered when needed [45]. Furthermore, it is advantageous when the material is nontoxic (after fabrication), applicable for post-processing, exhibits controllable or no volume changes and exhibits enough mechanical strength for the devised structure to be maintained after post-process.

Depending on the polymerization mechanism, or photosensitizers need to be added to the pre-polymer solution in order to enable the start of polymerization. Free radical generating photo-initiators are the most common (Table 21.2). They can also

Table 21.1 Material compilation

Material group	Chemical name	Reported method	Reference
Acrylic resin	TEGDM	Flow lithography	[43]
	WaterShed™ 11120	Microstereolithography	[45]
	ProtoTherm™ 12120	Microstereolithography	[45]
	Somos® 14120	Microstereolithography	[45]
	Acrylate resin R11	Microstereolithography	[13, 14, 46]
	Sartomer SR368, SR399, SR348	DLW	[47]
Thiol-ene resin	Pentaerythritol tetrakis (mercaptoacetate)	DLW	[44]
	PETE[a]	DLW	[44]
Negative photoresist	SU-8 2010	Microstereolithography	[34]
Positive photoresist	Microchem S-1813	Microstereolithography	[34]
Optical adhesives	NOA63	Microstereolithography	[34, 37]
	NOA89	Flow lithography	[43]
Organically modified ceramics	Ormocer/ormocomp[b]	DLW	[20, 48–51]
Hydrogels	PEG-DA[c]	Flow lithography	[16, 36, 43, 52–54]
	PEG-DA[c]	DLW	[20, 50, 55]
	PEG-DA[c]	*In situ* lithography	[19, 56]
	PETTA/PETA[d]	DLW	[20, 50, 51, 57]
	Biotin-PEG-acrylate	Flow lithography	[43]
	Acylamides	DLW	[58]
	BSA and other proteins	Microstereolithography	[22, 34, 35, 38, 59–61]
	Conjugated proteins (FITC-BSA, FITC-Avidin)	Microstereolithography	[34]
	BSA and other proteins	DLW	[62–64]
Soft polymers	PDMS[e]	DLW	[30]
Metal	AgNO$_3$/Ag	DLW	[65]

[a]Pentaerythritol tetraallyl ether
[b](–O–Si–O)
[c]Polyethylene(glycol) diacrylate
[d]Pentaerythritol tetraacrylate
[e]Poly(dimethylsiloxane)

be utilized as photosensitizers, enhancing the polymerization process. Similarly, diluents for viscosity control such as neopentyl glycol diacrylate (PNGD) [45] might also contribute to the polymerization process as so-called active diluent such as *n*-vinyl pyrrolidone [52]. In order to limit polymerization progression, inhibitors can be added to the pre-polymer solution such as hydroquinone [44]. To control curing depth, photo-absorbers are added to absorb light competitively to excitation events, e.g., SUDAN I photoabsorber [36] and Tinuvin 400 [45].

Table 21.2 List of common photo and photosensitizers

Photo-initiator/photosensitizer	References
DMPA (2,2-dimethoxy-2-phenylacetophenone)	[16, 43, 54]
Benzil	[58]
Darcour 1173 (2-hydroxy-2-methylpropiophenon)	[19, 43, 53]
DETC (7-diethylamino-3-thenoylcoumarin)	[44]
Eosin Y	[66, 67]
FAD (flavin adenine dinucleotide)	[22, 38, 60, 62]
Irgacure369 (2-benzyl-2-dimethylamino-1-(4-morpholinophenyl)-butanone-1)	[50, 55]
Irgacure 2959 (2-hydroxy-1-(4-(hydroxyethoxy) phenyl)-2-methyl-1-propane)	[52]
Isopropyl-thioxanthone	[30]
Lucirin TPO-L	[47]
Methylene blue	[60, 66, 68]
Omnirad 248 (2-benyl-2-(dimethylamino)-4′-morpholinobutyrophenone)	[58]
Phenylbis(2,4,6-trimethylbenzoyl)phosphine oxide	[36, 56]
Rose bengal	[59, 61, 63, 64]
RT dyes	[32]

Lastly, components are added to the pre-polymer solution such that functionality of these components is integrated into fabricated device. The maximum for feasible component concentrations depend on the components influence on transmissivity and polymerization reaction. For instance, in order to fabricate magnetic elements, superparamagnetic iron oxide (IONs) [43] or water-based ferrofluid EMG 508 [53] has been added. Similarly, for fluorescence imaging or particle coding, fluorescent beads such as FluoSpheres® [36] and polystyrene tracers have been added [16, 43]. Equally, in order to create conductive device elements, carbon nanotubes [69], graphene oxides [70] or ionic liquid [71] can be added. Using biocompatible hydrogels and avoiding long-term UV exposure, inclusion of cells during the polymerization enables cell-laden particles for further assembly [15, 52].

2.4 Polymerization Mechanism

The key point of photo-induced polymerization is significant progression of reaction in order to reach sustainable polymerization simultaneous to limiting or competitive reactions preventing polymerization of the entire pre-polymer amount.

There are multiple reaction mechanisms available, including acid generation where organic components become acidic upon photoexcitation and soluble to aqueous base developer [72] and hydrosilylation where platinum colloid acts as catalyst after UV decomposition [30], but by far the most common mechanism is radical chain reaction with multiple pathways for polymerization [32, 59, 73, 74]: (1) initiation, where radicals are generated via excitation, (2) propagation, where

radicals are generated simultaneously while polymer length grows via interaction of radicals and monomers, and (3) termination, where radicals are reacting with various partners without further propagation. Free radical initiated polymerization is usually exothermic reaction [75].

Excitation inside the focal volume of multiphoton excitation leads to the polymerization of a volume element, a voxel [29, 76]: For most polymerizations, the volume of polymerization represents the volume of laser focus, which can be described as an ellipsoid with diffraction limits in axial and radial direction as diameter and height. The voxel aspect ratio of height over diameter helps to characterize a voxel. Incomplete polymerization removed during post-processes [76] or diffusion processes [57] contribute to changes between the light volume shape and polymerized voxel shape. Understanding the effective voxel dimensions is crucial to guarantee sufficient connectivity between polymerized elements.

In order to predict voxel dimensions, required dose, i.e., laser power multiplied with irradiation time as absolute or area-relative or volume-relative value, and desired fabrication results, modeling and simulating the polymerization is essential. Generally, polymerization kinetics is assumed and then laser beam properties used to describe the translation from beam shape to polymerization. These models are separated by materials because it is depending on the type of reaction mechanism how the polymerization propagates through the material, i.e., by catalyst, by ion or by radical, depending on the material viscosity how diffusion influences propagating molecules, and it is depending on how radicals or ions are generated, i.e., as free radical, by depletion of the photo-initiator or monomer-bound radical: resins [77], metal [65], polymer [76], protein [62] and photoresists [48, 78–80]. Discrepancies between models and experimental results are often due to strong simplification of kinetic reactions and neglect of diffusion as suggested by experimental results [57]. Either infinitesimally small time spots are evaluated by simulation [76] or scaling factors compromise this discrepancy [62, 79].

2.5 Post-processing: Functionalization

Apart from dissolving unpolymerized pre-polymer materials and further causing polymerization by heat, UV or chemical treatment, there have been reports about post-process functionalization.

The goal of the functionalization varies with the device application: (1) It might be a goal to coat the polymer device with other surface material, such as metal coating [24, 81] and ceramic sintering [82], (2) it might be a goal to change the overall appearance of a structure from wireframe to voluminous shape where material deposition such as parylene-C chemical vapor deposition is a faster process substituting for photo-induced fabrication as soon as applicable [83], and (3) utilization of surface reactivity for Michael-Addition via malmeide, devices can be locally coated with metal [47], fluorescent groups [44] or with a light-sensitive group (e.g., MalPOEGMA) such that surface reactivity is controlled locally well-defined by laser excitation on demand [49].

3   Photo-Induced Fabrication Technologies for 3D Devices

The technologies examined in the following maintain the element of photoreactive material but vary the delivery of material and exposure of light into the material. These technologies were partially developed in parallel in pursuit of reducing fabrication time, reducing fabrication feature sizes, and enabling true 3D features.

3.1   Microstereolithography

Derived from stereolithography and rapid prototyping technologies, microstereolithography is a technology of bottom-up layer-stacking of micrometer-sized feature sizes [41]. A resolution of micrometer size became possible because photo-induced polymerization of resins was optimized with regards to laser technology making two-photon fabrication accessible and with regards to chemical advances enhancing photo-initiator efficiency [84].

The experimental setup of layer-stacking commonly is inverted with the light source placed below a glass container holding bulk pre-polymer solution in order to maximize fabrication stability, minimize viscous reflow effects and washing effort (Fig. 21.5). Often the light source is a halogen UV lamp utilizing one-photon absorption for polymerization excitation. Inspired by rapid prototyping and driven by industrialization, multiple working principles have been established, which can be grouped into vector-based fabrication and integral-layer-based fabrication [41].

The conventional microstereolithography is a vector-by-vector microstereolithography. In this technology, focused light is exposed along a programmed vector into the resin. The pre-polymer resin may be flattened and physically constrained by a glass surface (constrained surface technique), or it may be held freely in its container because the resin viscosity is sufficiently low for levelling the pre-polymer between fabrication steps quickly (free surface technique) [41]. Integral microstereolithography processes utilize dynamic-mask technologies in order to polymerize layer-to-layer where integrating light intensity over larger volumes [23, 60, 85].

Fig. 21.5 Principle of microstereolithography. A bulk of photoreactive material is distinctly exposed to light. Usually an inverted setup is in use uplifting polymerized layers by an elevator-driver and providing material from a container with glass window during the photoinduction
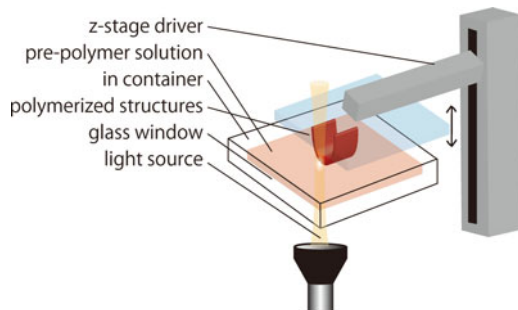
Post-processing of polymer structures fabricated by microstereolithography enables metal [81] and ceramic microfabrication [82]. With the development of a stage that supports pre-polymer solution exchange by syringe-pumps and reducing resin viscosity by diluent supplementation, fabrication of multimaterial microstructures was achieved [45].

The industrialization of microstereolithography is far advanced with various resins and devices available. Therefore, applications of this technology are not only widely distributed, but also bring the potential to be industrialized device elements themselves such as micro gears [86], wave guides, microtweezers or hearing aid caps [41]. Microfluidic devices that often harness laminar flow [3], fabricated by conventional photolithography have limited options for inlet position and upright orientation. Microstereolithography enables microfluidic devices with inlets of 3D position and orientation which is determined by device design and cannot be flexibly adjusted after fabrication. Dynamic-mask and layer-to-layer concept enable considerably faster fabrication speed for non-regular 3D devices than in conventional photolithography.

Here, we want to introduce 3D microfluidic fabrication devices applying microstereolithography to fabricate 3D elements (Fig. 21.6a–c), and *in situ* cell capture easily realized due to the high speed of integral-layer-to-layer fabrication (Fig. 21.6d–g). The embedment of liquids is axially aligned to the center, so-called coaxial alignment, in order to guarantee full embedment and to simplify device alignment. Such devices were readily fabricated by microstereolithography (Fig. 21.6a–c); Fig. 21.6a, b depicts an axisymmetric flow-focusing device (AFFD) embeds the flow of aqueous phase containing hydrogel polymer and cells into the flow of oily phase that can mass-fabricate monodisperse hydrogel beads [13, 46]. Figure 21.6c shows a double-coaxial laminar-flow microfluidic device which similarly to AFFD embeds axisymmetric core flow into two more sheath flows. Variation of flow rates and chemical constitution of the three inlet materials enables meter-long fiber fabrication. It is applied to various cell types provide hydrogel 3D culture condition and to enable handling for implantation [14]. Quick *in situ* fabrication enables single bacterial cell capture from multiple dispersed cells for studies of proliferation analysis of its progeny [23] (Fig. 21.6d–g).

When the desired 3D device is crucial to further fabrication steps and high-fabrication numbers might become relevant with progressed time; industrialization-optimized microstereolithography is a reliable fabrication technology.

3.2 *In Situ Photolithography*

In order to fabricate freely moving, but confined elements inside microfluidic devices, conventional photolithography requires many patterning steps, alignment, and removal of sacrificial material. *In situ* photolithography utilizes the concept of conventional photolithography with the keypoint that conventional photoresist is introduced into a target microfluidic device. The fabricated element is then freely
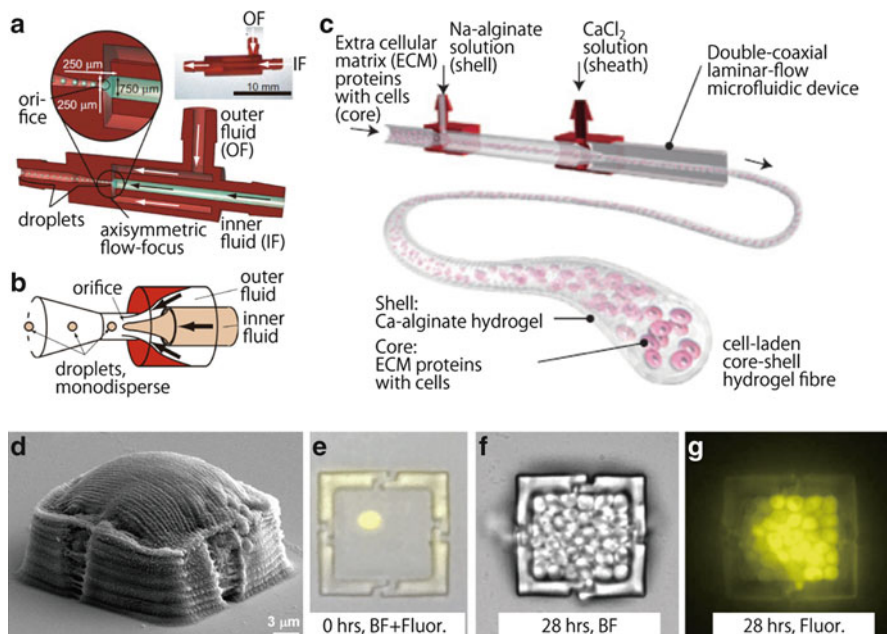
Fig. 21.6 Applications of microstereolithography for coaxial flow fabrication technologies. (**a**) Coaxial device for axisymmetric flow-focus of two different liquids, inlet as outer fluid and inner fluid, inset photograph shows device with scale bar of 10 mm [13]. (**b**) Oil and water phase form monodisperse droplets that have no contact with the channel walls [13]. (**c**) Two devices assembled with glass capillaries enable double coaxial laminar flow, three different liquids are guided for cell-laden fiber fabrication [14]. (**d, e**) A single yeast cell is isolated in a chamber, scale bar represents 3 μm, (**f, g**) with ongoing cell multiplication the progeny cells can be analyzed [23]. (**a, b**) Copyright 2009 Springer, reproduced (adapted) with permission from [13], (**c**) copyright 2013 Nature Publishing Group, reproduced with permission [14], (**d–g**) copyright 2012 American Chemical Society, reprinted (adapted) with permission from [23]

moving, but confined to its location inside the microfluidic device. For simple 3D elements, i.e., a laterally complex shape without in-depth complexity, the fabrication effort is considerably reduced in comparison with conventional lithography [19].

After introduction of photoresist into the target device, conventional fabrication steps are performed with the photoresist *in situ* inside the device (Fig. 21.7). Conventional photolithography masks, conventional alignment equipment and conventional post-processing such as photoresist development are sufficient for fabrication. Lateral resolution might be comparable to conventional results with dependency on photoresist, setup and light path through microfluidic components. Vertical resolution however is hypothesized to be dependent on microchannel thickness, light exposure duration and photo-absorption properties of the resists.

Application for this technology is broad as long as one can think of use for fabricated elements within fluidic devices, especially microfluidic PDMS devices. It is not required that the device is entirely freely moving. For instance, anchored gate
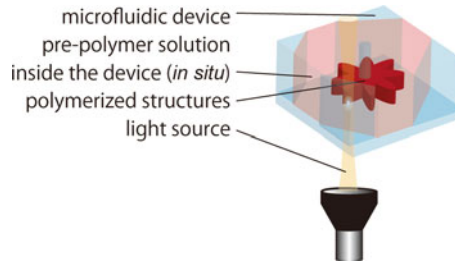
Fig. 21.7 Principle of *in situ* photolithography. A photoreactive material is introduced into the target microfluidic device. After alignment, a complex, free-moving microstructures is fabricated by light exposure into the photoreactive material
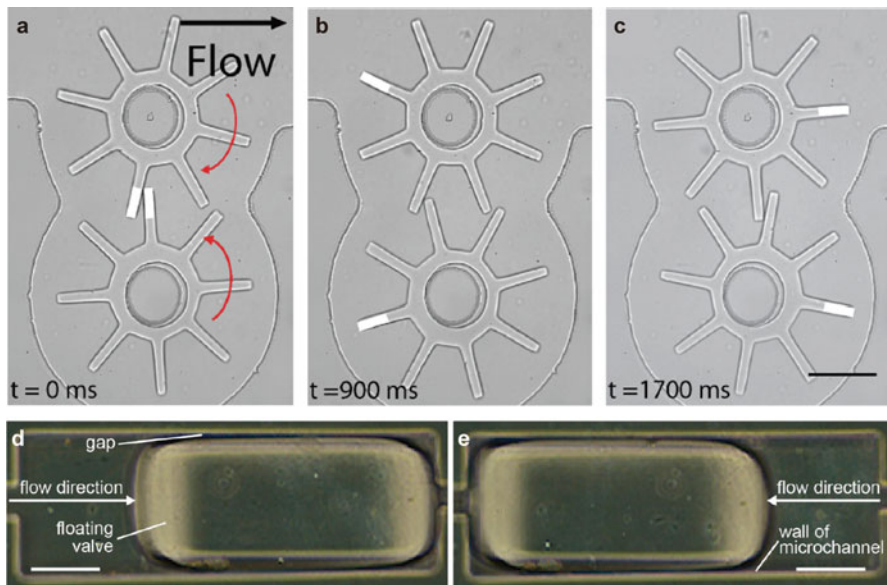
Fig. 21.8 Applications of *in situ* photolithography. (**a–c**) Microgear pair demonstrating co-rotation and flow-induced torque transmission [19]. Bright field images of a leaky valve [56]: a polymer block floats inside a microchannel according to flow direction, reducing released volume by clogging the exit in dependence of flow direction and applied pressure. (**a–c**) Copyright 2015 Springer, reproduced with permission from [19], (**d–e**) copyright 2016 by the authors, reproduced with permission from [56]. All scale bars indicate 100 μm

structures with flexible elements have been reported to passively regulate flow based on flow direction [18]. However, most reported applications focus on unanchored elements that provide function within the microfluidic device by moving [19, 56].

Rotary microgears have been fabricated around poles prepared within a microfluidic device (Fig. 21.8a–c) [19]. Such rotary microgears demonstrate rotation rate linearly controlled hydrodynamically and furthermore torque transmission between

paired microgears is documented. Such microgears can be employed for mixing within microfluidic devices, but also might be used to translate torque via microgears bound to pins into or from the microfluidic device for future application in microsurgery devices [19]. Pressure sensitive valves are devised by fabricating polymer blocks into cavities [56] (Fig. 21.8d,e); A polymer block moves with positive or negative pressure within its confining cavity, thereby opening or closing respective channels leading towards such cavity. By regulating the shape of such polymer block, flow rates are regulated in dependency of the applied pressures. By ensuring that valve and microfluidic device do not stick together, even at high pressures leakage at very low flow rates occurs [56]. Such low flow rates are problematic to achieve by conventional pump systems, but might find application in biological and medical studies.

In situ photolithography is a quick and reproducible technology for inserting movable structural elements into channel bearing devices, especially microfluidic devices.

3.3   Flow Lithography

Flow lithography uses the concept of microfluidic *in situ* photolithography to exploit the flow direction for continuous fabrication in order to achieve high-throughput of 3D microparticle fabrication [16, 43, 54]. Additionally, particle complexity is not only achieved by light patterning, but also utilizing inertial or laminar microfluidic liquid presentation [16]. Resolution can be 1–2 μm [87], but might fluctuate with device and photoreactive material.

Photoreactive material is flown inside a microfluidic device and light is exposed onto the channel (Fig. 21.9). Light exposure approaches one-step or single-shot particle formation for high-throughput when possible [15, 16, 43, 53, 54], but can rely on layer-to-layer formation [36, 52]. Currently, high-throughput and particle complexity are related competitively, resulting in a trade-off of mass fabrication
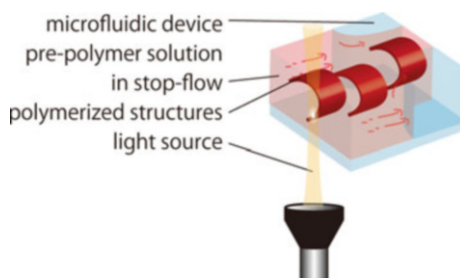
Fig. 21.9 Principle of flow lithography. A photoreactive material is introduced into a microfluidic device with channel design devised for fabrication. Complex, free-moving microstructures are fabricated by light exposure into the photoreactive material and then transported along the flow

for non-spherical particles. However, harnessing microfluidic effects can further enhance particle complexity without increase of fabrication steps; Changes in flow rate in combination with light exposure and device planning result in a broad range of fabrication strategies exploiting effects of laminar flow and inertial flow affected by obstacles and channel dimensions [16, 43]. Non-photoreactive material can be used to create voids. Addition of quenching chemicals and light absorbing chemicals to alter the polymerization process is another method to change the fabrication result. It might be exploited further that microfluidic channel dimensions can confine liquid to desired particle sizes. Confinement to the microfluidic channels dimension simultaneously represents a significant limitation for device fabrication. The field mainly focuses on rapid microparticle fabrication, but fiber fabrication with alteration of cross-sectional shape [36] is pursued to complement conventional fiber spinning technologies.

Four different technological subtypes can be distinguished. (1) Continuous flow lithography (CFL) is fabrication under continuous flow condition in a conventionally horizontal microfluidic device [16, 43, 53, 54]. In comparison, the continuous flow enables high-throughput, but remains limited in particle shape complexity even when utilizing laminar and inertial flow (Fig. 21.10a). (2) Stop-flow lithography (SFL) is fabrication under a condition in a conventionally horizontal microfluidic device where flow rate is altered to regulate exposure times and add layers [52]. In comparison, stop-flow lithography still has quite high-throughput especially when sequential steps are executed in parallel or multifocal excitation is used, particle complexity can be increased while still being simple to execute. (3) Lock-and-release lithography (LRL) is fabrication of continuous flow or stop-flow condition within a microfluidic channel that utilizes PDMS deformation to hold or release particles [15]. Its disadvantage relying on the PDMS deformation by a second bonded channel, it requires alignment and computer-assisted synchronization. The PDMS device opposed limitation to particle shape is simultaneously disadvantage and advantage; the shapes are specific to the PDMS device, but the PDMS device might be devised to separate or guide these specific shapes. (4) Vertical flow lithography (VFL) is fabrication of continuous flow or stop-flow condition in a vertical microfluidic channel which aligns light exposure path and flow direction [36] (Fig. 21.10b–d). While vertical flow might be challenging to operate and permits only one fabrication localization per device, this technology has demonstrated a wide range of multilayered and tapered anisotropic shapes, and angular compartmental material combinations, including cross-sectional changes in fiber elements.

Because materials and shapes can be varied greatly, applications vary similarly. Microparticles for assembly have been suggested using lock-key shape recognition [16], trail-assisted assembly [15] or magnetic control [43, 53]. Optical tools such as biconcave lenses have reported [88]. Tissue engineering or cell culture studies can be approached with cell-laden hydrogel microparticles [15, 52], but should best be fabricated with other than the conventionally used UV-light. Flow lithography is generally perceived to find application in drug delivery and self-assembly [36, 87, 89].
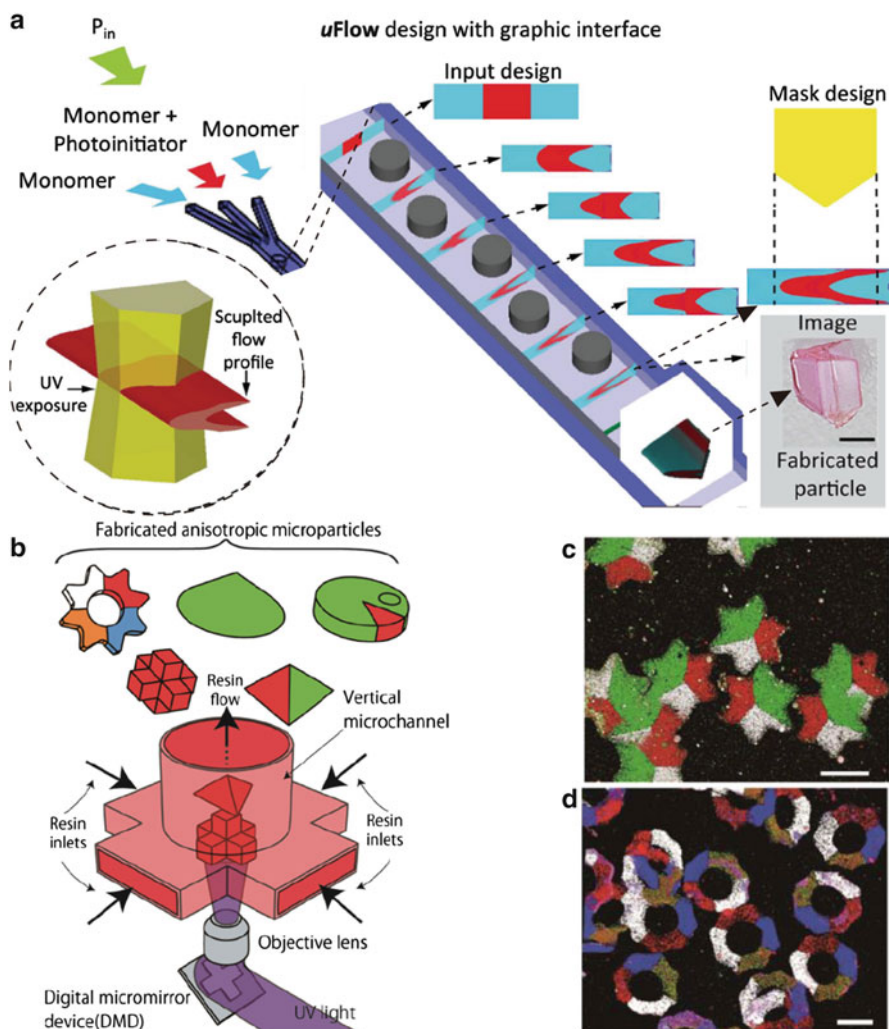
Fig. 21.10 Applications of flow lithography of microparticle fabrication for (**a**) assembly and (**b–d**) drug delivery or cell culture. (**a**) Laminar flow profile changes due to inertial flow reacting to channel obstacles, a single-shot shape is exposed at polymerization sites with desired profiles [43]. (**b**) Vertical flow lithography concept with four different resin inlets [36]. (**c, d**) Three or four inlets are used with differently colored hydrogel during exposure of (**c**) a *star* shape and (**d**) a *ring* shape. (**a**) Copyright 2015 John Wiley and Sons reproduced (adapted) with permission from [43], (**b–d**) copyright 2016 John Wiley and Sons reproduced (adapted) with permission from [36]

Fig. 21.11 Principle of direct laser writing (DLW). A laser is focused into the photoreactive material directly. Computer-guided arbitrary movement of the laser focus translates into correspondingly polymerized arbitrary 3D structures

pre-polymer solution
polymerized structures
pulsed laser light
stage control
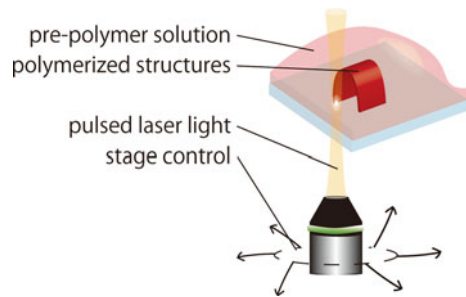
3.4 Direct Laser Writing

In direct laser writing (DLW), the focus of a laser with high peak intensities and short pulse duration is exposed to photoreactive material directly [1, 29]. Direct laser writing causes polymerization initiated along an arbitrary path of the laser focus through the pre-polymer solution. Operation with a setup that harnesses two-photon or three-photon excitation, the polymerization is confined to the laser focus volume and submicron resolution is achieved [29, 48, 62]. Submicron resolution and free arbitrary fabrication path are the advantages of this technology.

Direct laser writing setup usually utilizes a large wavelength (e.g., 780 nm) laser with femtosecond (e.g., 100 fs) pulse width in order to polymerize materials that are sensitive to UV-light (e.g., 390 nm) (Fig. 21.11). Such setup choice harnesses on the one hand that large wavelength light easily travels undisturbedly though material, and on the other hand that enormous research efforts and industrialization have driven the development of UV-sensitive materials as previously illustrated. Pulsed laser light bears the advantage that photon intensity is temporally and spatially confined, enhancing two-photon absorption, simultaneously reducing laser conveyed heat in comparison to continuous-wave laser light. The objective guiding the laser into the photoreactive material is usually mounted on a piezo-actuated stage. This conventional setup is physically limited when the lens reaches the glass substrate, but normally already is challenged by the diffusion of the laser focus increasing due to diffraction through pre-polymer solution.

In order to achieve high-aspect ratio structures with consistent feature sizes regardless the distance to substrate surface and substrate transmissivity, Dip-in-Lens Lithography (DiLL) was developed [25] (Fig. 21.12). By immersing the lens into the photoreactive material instead of immersion oil, movement limitations are removed and a constant lens to fabrication area guaranteed. It requires the photoreactive material to have a refractive index matching the lens in order to achieve small feature sizes. Furthermore, the technology is sensitive to quickly losing lens quality by remaining photoresists due to sloppy cleaning or accidental white light exposure.

Fabricating micro-to-mesoscale devices while controlling nanometer-sized feature sizes is a driving force of DLW growing applications ranging already from metamaterials, proteinaceous microstructures, pH-actuated devices to well-defined

Fig. 21.12 Dip-in-lens
lithography (DiLL) for DLW.
The lens is immersed into the
photoreactive material during
the DLW fabrication process.
The lens moves away from
the substrate with a constant
distance to fabrication area
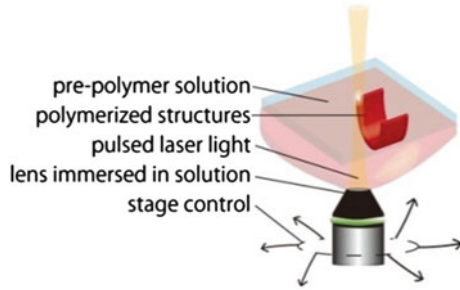
3D microenvironments (Fig. 21.13). Because DLW enables structuring with sub-
micron feature sizes, fabrication of metamaterials for the expansion of material
properties beyond the natural possible [27, 28] was recently realized for optical
metamaterials [24] and mechanical metamaterials [25, 26] (Fig. 21.13a, b). In
addition, wireframe microstructures as rings have been explored for wallless liquid
pathways [90] and for proteinaceous device [62] (Fig. 21.13c). Polymerization of
hydrogels controlling the cross-linking density by exposure parameters has given
rise to pH-sensitive soft-microrobots [58, 63, 91] (Fig. 21.13d, e) and tunable arrays
[68, 92]. Cell culture studies were reportedly performed on 3D sequential multi-
component-fabricated scaffolds to study interaction of 3D environment and cell
morphology [20, 50] (Fig. 21.13f, g), proteinaceous scaffolds to study cell migration
[64] or pillar arrays for force measurements [55].

   DLW technologies approach and even succeed the classical diffraction limits, as
well as offer fabrication of truly 3D structure features. The price for such precision
is a considerably slower fabrication time than previously introduced technologies.
Already, specialized mirror systems (Galvano technology) and parallelized fabrica-
tion by beam splitting are in use.

3.5 *Overviewing Summary*

Table 21.3 summarizes advantages and disadvantages of the previously reviewed
fabrication technologies symbolizing degree of advantage by one, two or three plus
signs '+' and disadvantage by a minus sign '−'.

4 Conclusion

In this chapter, we introduced fundamental concepts of photo-induced fabrication
for 3D microdevices. Complementary to conventional photolithography, 3D device
fabrication is required to approach conventionally challenging shapes and freeforms.
While being complementary in approach and utilization, photo-induced fabrication
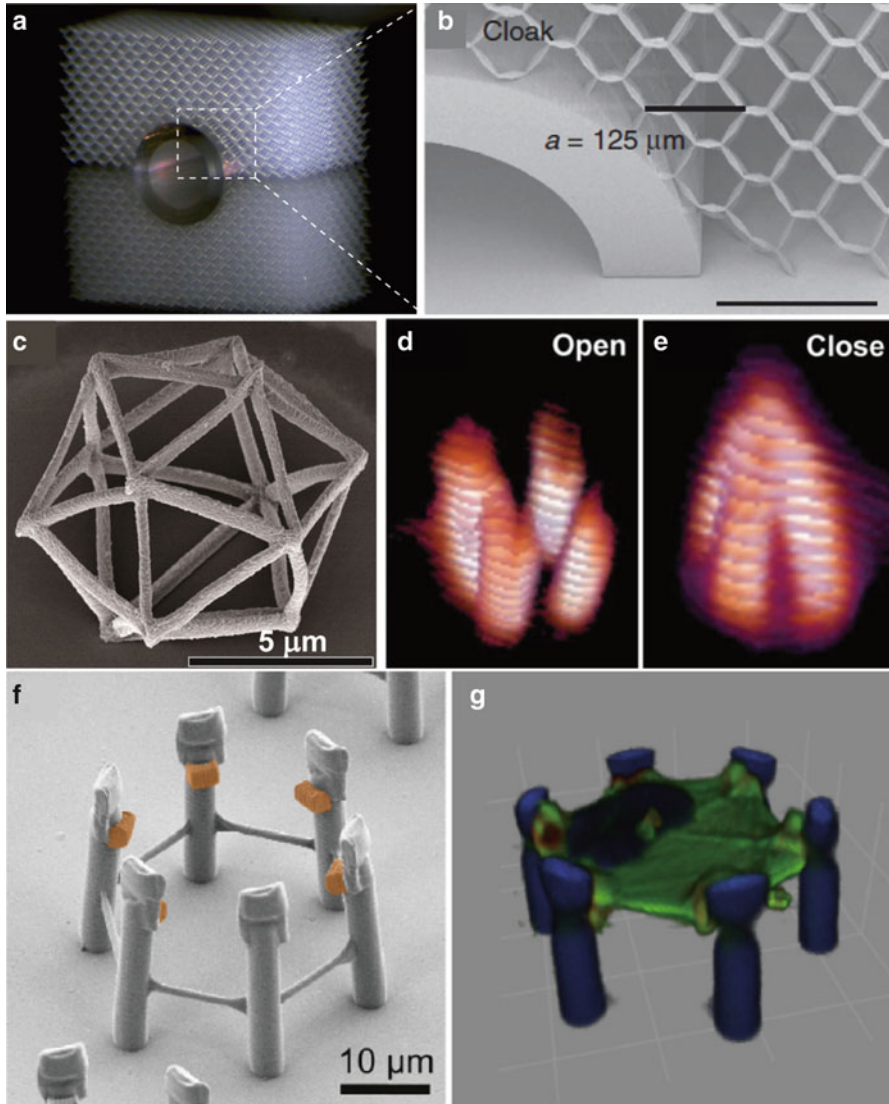
Fig. 21.13 Applications of DLW. (**a**) Cloaking metamaterial photograph, with (**b**) SEM image of the network [26]. (**c**) Proteinaceous wireframe microparticle [62]. (**d, e**) Proteinaceous 3D microgripper in pH-actuated (**d**) open and (**e**) closed state [63]. (**f, g**) Two-component cell culture scaffold for subcellular stretching of mammalian cells; (**f**) false-colored SEM image and (**g**) 3D reconstruction by laser scanning confocal microscopy of scaffold with cells [50]. Reproduced (adapted) with permission (**a, b**) Copyright 2014 Nature Publishing Group, from [26], (**c**) Copyright 2015 AIP Publishing LLC from [62], (**d, e**) Copyright 2015 John Wiley and Sons from [63], (**f, g**) Copyright 2015 Elsevier from [50]

Table 21.3 Advantage and disadvantage of each fabrication technology

	Microstereolithography	*In situ* photolithography	Flow lithography	Direct laser writing
Subsection	3.1	3.2	3.3	3.4
High-throughput	+++	−	++	−
Integrated device	+	+++	−	++
Speed	++	+++	+++	+
Resolution	+	+	++	+++
True 3D	++	+	+	+++

Advantage and degree of advantage are symbolized by plus signs "+" and disadvantage is symbolized by a minus sign "−"

for 3D microdevices harnesses advances in material and tool development from extensive research in the field of conventional photolithography.

We reviewed four technologies that distinguish in light path and presentation of photoreactive material to the exposure. Each of the presented technologies has advantages and disadvantages, which enables very specific device fabrication conditions for broad and overlapping applications ranging from microfluidics over bioMEMS to cell culture studies. The field of photo-induced fabrication is still strongly developing striving to quickly fabricate well-controlled devices with small feature sizes.

References

1. Maruo S, Fourkas JT (2008) Recent progress in multiphoton microfabrication. Laser Photonics Rev 2:100–111
2. Vaezi M, Seitz H, Yang S (2013) A review on 3D micro-additive manufacturing technologies. Int J Adv Manuf Technol 67:1721–1754
3. Whitesides GM (2006) The origins and the future of microfluidics. Nature 442:368–373
4. Nakamaru Y, Honma H (2009) Fabrication of three-dimensional microstructure by nickel plating and photolithography. Trans Inst Met Finish 87:259–263
5. Larramendy F, Serien D, Yoshida S, Jalabert L, Takeuchi S, Paul O (2015) High-topography surface functionalization based on parylene-C peel-off for patterned cell growth. In: MEMS 2015 Proc., pp 328–331
6. Däschner W, Long P, Larsson M, Lee SH (1995) Fabrication of diffractive optical elements using a single optical exposure with a gray level mask. J Vac Sci Technol B 13:2729–2731
7. Hayashi T, Shibata T, Kawashima T, Makino E, Mineta T, Masuzawa T (2008) Photolithography system with liquid crystal display as active gray-tone mask for 3D structuring of photoresist. Sensors Actuators A 144:381–388
8. Miyake M, Chen Y-C, Braun PV, Wiltzius P (2009) Fabrication of three-dimensional photonic crystals using multibeam interference lithography and electrodeposition. Adv Mater 21:3012–3015
9. Suslik L, Pudis D, Skriniarova J, Martincek I, Kubicova I, Kovac J (2012) 2D photonic structures for optoelectronic devices prepared by interference lithography. Phys Procedia 32:807–813

10. Hirai Y, Inamoto Y, Sugano K, Tsuchiya T, Tabata O (2007) Moving mask UV lithography for three-dimensional structuring. J Micromech Microeng 17:199–206

11. Larramendy F, Mazenq L, Temple-Boyer P, Nicu L (2012) Three-dimensional closed microfluidic channel fabrication by stepper projection single step lithography: the diabolo effect. Lab Chip 12:387–390

12. Maruo S, Inoue H (2006) Optically driven micropump produced by three-dimensional two-photon microfabrication. Appl Phys Lett 89:144101

13. Morimoto Y, Tan W-H, Takeuchi S (2009) Three-dimensional axisymmetric flow-focusing device using stereolithography. Biomed Microdevices 11:369–377

14. Onoe H, Okitsu T, Itou A, Kato-Negishi M, Gojo R, Kiriya D, Sato K, Miura S, Iwanaga S, Kuribayashi-Shigetomi K, Matsunaga YT, Shimoyama Y, Takeuchi S (2013) Metre-long cell-laden microfibres exhibit tissue morphologies and functions. Nat Mater 12:584–590

15. Chung SE, Park W, Shin S, Lee SA, Kwon S (2008) Guided and fluidic self-assembly of microstructures using railed microfluidic channels. Nat Mater 7:581–587

16. Paulsen KS, Di Carlo D, Chung AJ (2015) Optofluidic fabrication for 3D-shaped particles. Nat Commun 6:6976

17. Di Carlo D (2009) Inertial microfluidics. Lab Chip 9:3038–3046

18. Sochol RD, Glick CC, Lee KY, Brubaker T, Lu A, Wah M, Gao S, Hicks E, Wolf KT, Iwai K, Lee LP, Lin L (2013) Single-layer "domino" diodes via optofluidic lithography for ultra-low reynolds number applications. In: MEMS 2013 Proc., pp 153–156

19. Moon B-U, Tsai SSH, Hwang DK (2015) Rotary polymer micromachines: *in situ* fabrication of microgear components in microchannels. Microfluid Nanofluid 19:67–74

20. Klein F, Richter B, Striebel T, Franz CM, vonFreymann G, Wegener M, Bastmeyer M (2011) Two-component polymer scaffolds for controlled three-dimensional cell culture. Adv Mater 23:1341–1345

21. Kim S, Qiu F, Kim S, Ghanbari A, Moon C, Zhang L, Nelson BJ, Choi H (2013) Fabrication and characterization of magnetic microrobots. Adv Mater 25:5863–5868

22. Kaehr B, Allen R, Javier DJ, Currie J, Shear JB (2004) Guiding neuronal development with *in situ* microfabrication. Proc Natl Acad Sci USA 101:16104–16108

23. Harper JC, Brozik SM, Brinker CJ, Kaehr B (2012) Biocompatible microfabrication of 3D isolation chambers for targeted confinement of individual cells and their progeny. Anal Chem 84:8985–8989

24. Rill MS, Plet C, Thiel M, Staude I, vonFreymann G, Linden S, Wegener M (2008) Photonic metamaterials by direct laser writing and silver chemical vapour deposition. Nat Mater 7:543–546

25. Bückmann T, Stender N, Kadic M, Kaschke J, Frölich A, Kennerknecht T, Eberl C, Thiel M, Wegener M (2012) Tailored 3D mechanical metamaterials made by dip-in direct-laser-writing optical lithography. Adv Mater 24:2710–2714

26. Bückmann T, Thiel M, Kadic M, Schittny R, Wegener M (2014) An elasto-mechanical unfeelability cloak made of pentamode metamaterials. Nat Commun 5:4130

27. Milton GW, Cherkaev AV (1995) Which elasticity tensors are realizable? J Eng Mater Technol 117:483–493

28. Kadic M, Bückmann T, Schittny R, Wegener M (2013) Metamaterials beyond electromagnetism. Rep Prog Phys 76:126501

29. Fischer J, Wegener M (2013) Three-dimensional optical laser lithography beyond the diffraction limit. Laser Photonics Rev 7:22–44

30. Coenjarts CA, Ober CK (2004) Two-photon three-dimensional microfabrication of poly(dimethylsiloxane) elastomers. Chem Mater 16:5556–5558

31. Bhawalkar JD, He GS, Prasad PN (1996) Nonlinear multiphoton processes in organic and polymeric materials. Rep Prog Phys 59:1041–1070

32. Lee K-S, Kim RH, Yang D-Y, Park SH (2008) Advances in 3D nano/microfabrication using two-photon initiated polymerization. Prog Polym Sci 33:631–681

33. Chan KF, Feng Z, Yang FR, Ishikawa A, Mei W (2003) High-resolution maskless lithography. J Microlith Microfab Microsyst 2:331–339

34. Jennes NJ, Hill RT, Hucknall A, Chilkoti A, Clark RL (2010) A versatile diffractive maskless lithography for single-shot and serial microfabrication. Opt Express 18:11754–11762
35. Nielson R, Kaehr B, Shear JB (2009) Microreplication and design of biological architectures using dynamic-mask multiphoton lithography. Small 5:120–125
36. Habasaki S, Lee WC, Yoshida S, Takeuchi S (2016) Vertical flow lithography for fabrication of 3D anisotropic particles. Small 11:6391–6396
37. Jennes NJ, Wulff KD, Johannes MS, Padgett MJ, Cole DG, Clark RL (2008) Three-dimensional parallel holographic micropatterning using a spatial light modulator. Opt Express 16:15942–15948
38. Lawson JL, Jenness N, Wilson S, Clark RL (2013) Method of creating microscale prototypes using SLM based holographic lithography. SPIE 8612(86120L):1–8
39. Bertsch A, Jézéquel JY, André JC (1997) Study of the spatial resolution of a new 3D microfabrication process: the microstereophotolithography using a dynamic mask-generator technique. J Photochem Photobiol A Chem 107:275–281
40. Ritschdorff ET, Nielson R, Shear JB (2012) Multi-focal multiphoton lithography. Lab Chip 12:867–871
41. Bertsch A, Bernhard P, Renaud P (2001) Microstereolithography: concepts and applications. In: Emerging technologies and factory automation proceedings, vol 2, pp 289–298
42. Tormen M, Businaro L, Altissimo M, Romanato F, Cabrini S, Perennes F, Proietti R, Sun H-B, Kawata S, DiFabrizio E (2004) 3D patterning by means of nanoimprinting, X-ray and two-photon lithography. Microelectron Eng 73–74:535–541
43. Wu C-Y, Owsley K, Di Carlo D (2015) Rapid software-based design and optical transient liquid molding of microparticles. Adv Mater 27:7970–7978
44. Quick AS, Fischer J, Richter B, Pauloehrl T, Trouillet V, Wegener M, Barner-Kowollik C (2013) Preparation of reactive three-dimensional microstructures via direct laser writing and thiol-ene chemistry. Macromol Rapid Commun 34:335–340
45. Choi J-W, MacDonald E, Wicker R (2010) Multi-material microstereolithography. Int J Adv Manuf Technol 49:543–551
46. Matsunaga YT, Morimoto Y, Takeuchi S (2010) Monodisperse cell-encapsulating peptide microgel beads for 3D cell culture. Langmuir 26:2645–2649
47. Farrer RA, LaFratta CN, Li L, Praino J, Naughton MJ, Saleh BEA, Teich MC, Fourkas JT (2006) Selective functionalization of 3-D polymer microstructures. J Am Chem Soc 128:1796–1797
48. Serbin J, Egbert A, Ostendorf A, Chichkov BN, Houbertz R, Domann G, Schulz J, Cronauer C, Fröhlich L, Popall M (2003) Femtosecond laser-induced two-photon polymerization of inorganic–organic hybrid materials for applications in photonics. Opt Lett 28:301–303
49. Richter B, Pauloehrl T, Kaschke J, Fichtner D, Fischer J, Greiner AM, Wedlich D, Wegener M, Delaittre G, Barner-Kowollik C, Bastmeyer M (2013) Three-dimensional microscaffolds exhibiting spatially resolved surface chemistry. Adv Mater 25:6117–6122
50. Scheiwe AC, Frank SC, Autenrieth TJ, Bastmeyer M, Wegener M (2015) Subcellular stretch-induced cytoskeletal response of single fibroblasts within 3D designer scaffolds. Biomaterials 44:186–194
51. Greiner AM, Klein F, Gudzenko T, Richter B, Striebel T, Wundari BG, Autenrieth TJ, Wegener M, Franz CM, Bastmeyer M (2015) Cell type-specific adaptation of cellular and nuclear volume in micro-engineered 3D environments. Biomaterials 69:121–132
52. Panda P, Ali S, Lo E, Chung BG, Hatton TA, Khademhosseini A, Doyle PS (2008) Stop-flow lithography to generate cell-laden microgel particles. Lab Chip 8:1056–1061
53. Hakimi N, Tsai SSH, Cheng C-H, Hwang DK (2014) One-step two-dimensional microfluidics-based synthesis of three-dimensional particles. Adv Mater 26:1393–1398
54. Chung SE, Park W, Park H, Yu K, Park N, Kwon S (2007) Optofluidic maskless lithography system for real-time synthesis of photopolymerized microstructures in microfluidic channels. Appl Phys Lett 91:041106

55. Sochol RD, Heo YJ, Iwanaga S, Lei J, Wolf KT, Lu A, Kurihara M, Mori S, Serien D, Song L, Lin L, Takeuchi S (2013) Cells on arrays of microsprings: an approach to achieve triaxial control of substrate stiffness. In: MEMS 2013 Proc., pp 90–93

56. Morimoto Y, Mukouyama Y, Habasaki S, Takeuchi S (2016) Balloon Pump with Floating Valves for Portable Liquid Delivery. Micromachines 7:39

57. Mueller JB, Fischer J, Mayer F, Kadic M, Wegener M (2014) Polymerization kinetics in three-dimensional direct laser writing. Adv Mater 26:6566–6571

58. Xiong Z, Zheng M-L, Dong X-Z, Chen W-Q, Jin F, Zhao Z-S, Duan X-M (2011) Asymmetric microstructure of hydrogel: two-photon microfabrication and stimuli-responsive behavior. Soft Matter 7:10353–10359

59. Pitts JD, Campagnola PJ, Epling GA, Goodman SL (2000) Submicron multiphoton free-form fabrication of proteins and polymers: studies of reaction efficiencies and applications in sustained release. Macromolecules 33:1514–1523

60. Kaehr B, Jason BS (2008) Multiphoton fabrication of chemically responsive protein hydrogels for microactuation. Proc Natl Acad Sci USA 105:8850–8854

61. Basu S, Campagnola PJ (2004) Enzymatic activity of alkaline phosphatase inside protein and polymer structures fabricated via multiphoton excitation. Biomacromolecules 5:572–579

62. Serien D, Takeuchi S (2015) Fabrication of submicron proteinaceous structures by direct laser writing. Appl Phys Lett 107:013702

63. Lee MR, Phang IY, Cui Y, Lee YH, Ling XY (2015) Shape-shifting 3D protein microstructures with programmable directionality via quantitative nanoscale stiffness modulation. Small 11:740–748

64. Sie YD, Li Y-C, Chang N-S, Campagnola PJ, Chen S-J (2015) Fabrication of three-dimensional multi-protein microstructures for cell migration and adhesion enhancement. Biomed Opt Express 6:480–490

65. Tanaka T, Ishikawa A, Kawata S (2006) Two-photon-induced reduction of metal ions for fabricating three-dimensional electrically conductive metallic microstructure. Appl Phys Lett 88:081107

66. Spivey EC, Ritschdorff ET, Connell JL, McLennon CA, Schmidt CE, Shear JB (2013) Multiphoton lithography of unconstrained three-dimensional protein microstructures. Adv Funct Mater 23:333–339

67. Iosin M, Stephan O, Astilean S, Duperray A, Baldeck PL (2007) Microstructuration of protein matrices by laser-induced photochemistry. J Optoelectron Adv Mater 9:716–720

68. Sun Y-L, Dong W-F, Niu L-G, Jiang T, Liu D-X, Zhang L, Wang Y-S, Chen Q-D, Kim D-P, Sun H-B (2014) Protein-based soft micro-optics fabricated by femtosecond laser direct writing. Light Sci Appl 3, e129

69. Guo Q, Xiao S, Aumann A, Jaeger M, Chakif M'B, Ghadiri R, Esen C, Ma M, Ostendorf A (2012) Using laser microfabrication to write conductive polymer/SWNTs nanocomposites. J Laser Micro/Nanoeng 7:44–48

70. Lin C-Y, Hsu K-M, Huang H-C, Yeh T-F, Chang H-Y, Lien C-H, Teng H, Chen S-J (2015) Multiphoton fabrication of freeform polymer microstructures containing graphene oxide and reduced graphene oxide nanosheets. Opt Mater Express 5:218–226

71. Bakhtina NA, Voigt A, MacKinnon N, Ahrens G, Gruetzner G, Korvink JG (2015) Novel ionic liquid—polymer composite and an approach for its patterning by conventional photolithography. In: MEMS 2015 Proc., pp 97–101

72. Yu T, Ober CK, Kuebler SM, Zhou W, Marder SR, Perry JW (2003) Chemically amplified positive resists for two-photon three-dimensional microfabrication. Adv Mater 15:517–521

73. Maruo S, Nakamura O, Kawata S (1997) Three-dimensional microfabrication with two-photon-absorbed photopolymerization. Opt Lett 22:132–134

74. Fouassier JP, Morlet-Savary F, Lalevée J, Allonas X, Ley C (2010) Dyes as photoinitiators or photosensitizers of polymerization reactions. Materials 3:5130–5142

75. Oster G (1954) Dye-sensitized photopolymerization. Nature 173:300–301

76. Uppal N, Shiakolas PS (2008) Modeling of temperature-dependent diffusion and polymerization kinetics and their effects on two-photon polymerization dynamics. J Micro/Nanolithogr MEMS MOEMS 7:043002
77. Wang I, Bouriau M, Baldeck PL, Martineau C, Andraud C (2002) Three-dimensional microfabrication by two-photon-initiated polymerization with a low-cost microlaser. Opt Lett 27:1348–1350
78. Sun H-B, Tanaka T, Kawata S (2002) Three-dimensional focal spots related to two-photon excitation. Appl Phys Lett 80:3673–3675
79. Sun H-B, Takada K, Kim M-S, Lee K-S, Kawata S (2003) Scaling laws of voxels in two-photon photopolymerization nanofabrication. Appl Phys Lett 83:1104–1106
80. Lim TW, Park SH, Yang DY, Kong HJ, Lee KS (2006) Direct single-layered fabrication of 3D concavo–convex patterns in nano-stereolithography. Appl Phys A 84:379–383
81. Ikuta K, Hirowatari K (1993) Real three dimensional micro fabrication using stereo lithography and metal molding. In: MEMS'93 Proc., pp 42–47
82. Bertsch A, Sebastien J, Renaud P (2004) Microfabrication of ceramic components by microstereolithography. J Micromech Microeng 14:197–203
83. Kurihara M, Heo YJ, Kuribayashi-Shigetomi K, Takeuchi S (2012) 3D laser lithography combined with parylene coating for the rapid fabrication of 3D microstructures. In: MEMS 2012 Proc., pp 196–199
84. Cumpston BH, Ananthavel SP, Barlow S, Dyer DL, Ehrlich JE, Erskine LL, Heikal AA, Kuebler SM, Lee I-YS, McCord-Maughon D, Qin J, Roeckel H, Rumi M, Wu X-L, Marder SR, Perry JW (1999) Two-photon polymerization initiators for three-dimensional optical data storage and microfabrication. Nature 398:51–54
85. Limaye AS, Rosen DW (2007) Process planning method for mask projection microstereolithography. Rapid Prototyp J 13:76–84
86. Ikuta K, Maruo S, Kojima S (1998) New micro stereolithography for freely movable 3D micro structure—SuperIH process with submicron resolution. In: MEMS'98 Proc., pp 290–295
87. Dendukuri D, Doyle PS (2009) The synthesis and assembly of polymeric microparticles using microfluidics. Adv Mater 21:4071–4086
88. Nisisako T, Ando T, Hatsuzawa T (2014) Capillary-assisted fabrication of biconcave polymeric microlenses from microfluidic ternary emulsion droplets. Small 10:5116–5125
89. Shepherd RF, Conrad JC, Rhodes SK, Link DR, Marquez M, Weitz DA, Lewis JA (2006) Microfluidic assembly of homogeneous and Janus colloid-filled hydrogel granules. Langmuir 22:8618–8622
90. Lee WC, Heo YJ, Takeuchi S (2012) Wall-less liquid pathways formed with 3-dimensional microring arrays. Appl Phys Lett 101:114108
91. Serien D, Takeuchi S (2015) Chemically responsive protein-photoresist hybrid actuator. In: MEMS 2015 Proc., pp 470–471
92. Lay CL, Lee MR, Lee HK, Phang IY, Ling XY (2015) Transformative two-dimensional array configurations by geometrical shape-shifting protein microstructures. ACS Nano 9:9708–9717

Chapter 22
Long-Range Nano-Scanning Devices Based on Optical Sensing Technology

ChaBum Lee

Abstract A compact, low-cost, and high-performance nanopositioning system for dimensional metrology tools. A polymer-based large-range nanopositioning system was fabricated by an additive manufacturing (AM) process, stereolithography. The AM flexural properties were characterized to design the nanopositioning stage capable of millimeter range motion. It was considered to be responsible for AM fabrication tolerance or local irregularity in material properties because those properties are highly dependent on curing temperature and time even though the stage is fabricated layer-by-layer under the identical condition. Also, a design principle of an optical knife-edge sensor (OKES) is to provide a long range and high accuracy for nanopositioning. The OKES utilizes interference fringe due to optical knife-edge diffraction and has more than 40 % of high sensor sensitivity slope. The OKES was modeled by using electromagnetic wave propagation principle and experimentally verified by testing the OKES with the *XY* stage. The sensor noise level shows *X* 20.1 nm and Y 19.4 nm, and the fundamental sensing limit of the OKES was estimated *X* 0.19 nm/$\sqrt{\text{Hz}}$ and *Y* 0.23 nm/$\sqrt{\text{Hz}}$ for ± 1.0 mm working range. In the *XY* circular motion, 0.2/1.0/2.0 mm diameter at 1 Hz, the trajectory root-mean-square errors were 0.77/2.50/3.50 μm, respectively. These results indicate that the OKES can be a good alternative to precision metrology tools in terms of large working range, positioning accuracy, resolution, linearity, bandwidth, and control effectiveness.

C. Lee (✉)
Department of Mechanical Engineering, Tennessee Technological University, Cookeville, TN, USA
e-mail: clee@tntech.edu

1   Introduction

Displacement sensing is a fundamental technique that is used in many fields to relate the positioning of two bodies in space or to combine features of size with imaging. For example, displacement sensors are used to determine the *X*- and *Y*-lateral positions of an atomic force microscopy (AFM) probe relative to the sample [1–3]. As the AFM probe is rastered across the sample, the image signal combined with the lateral locations can be used to build an image with known features of size. With a rapid progress in the field of nanotechnology, there is high demand for new displacement measurement technologies to be able to achieve sub-nanometer level resolution with long-term reliability. These techniques are commonly applied to the measurement tools of machine tools, pick and place machines, coordinate measuring machines and semiconductor manufacturing equipment [3–9].

Capacitive sensors (CS), linear variable differential transformers (LVDT), optical linear encoders, and laser interferometers are the most commonly used in precision metrology areas such as nano-scanning devices or instruments [2, 3]. As illustrated in Fig. 22.1, CS and LVDT are compact nanoscale sensors with a few hundredths of micrometer range. On the other hand, optical linear encoders and laser interferometers are large scale sensors with moderate-to-high accuracy intended for ranges over a few meters. There is however little documentation on mesoscale displacement sensing technology for the measuring range 100 μm–5 mm with high accuracy less than 50 nm with a low cost. Achieving a large displacement range (>1 mm) along with nanometer level displacement accuracy (<50 nm), simultaneously, has been a key challenge in nanopositioning systems due to physical limitations associated with the driving mechanisms, actuators, sensors, and their integration. Currently, many commercially available nanopositioning stages have been restricted to the range of approximately a few hundreds of micrometers per axis [10–15]. There is, however, now a growing market need for multi-axis nanopositioning systems that can provide a millimeter range while keeping the nanometer level dimensional accuracy in a compact integration and low cost.
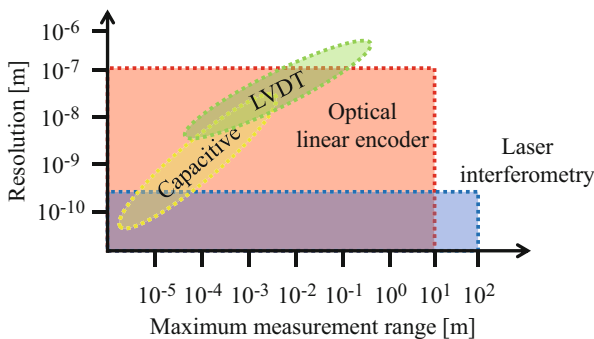
Fig. 22.1 Precision displacement sensors [2]

Recently, precision manufacturing technologies such as machine tools, semiconductor manufacturing processes, scanning probe microscopes, and coordinate measuring machines have been rapidly processed and stimulated [1, 3, 12]. The rapidly growing and ever-increasing applications of nanotechnology have increased the demand for high-speed and high-precision nanopositioning systems. Precision positioning stages play a key role of major breakthroughs in broad research areas and industrial applications including bionanotechnology, nanometrology, semiconductor manufacturing, chemical science and engineering, nanomachining and nanofabrication, material science, and high-density data storage systems [13–17]. Those positioning stages in general rely on flexures to provide linear or rotational motions with no friction or stiction. The emergence of compliant, piezoelectric stack-actuated nanopositioners fulfils the requirements of nanotechnology related applications [12]. There recently has been a trend away from costly conventional methods toward small-scale and cost-efficient prototyping and production facilities called microfactories [17–20]. Although there have been some impressive developments recently in these manufacturing techniques, they are still limited by their lack of affordability and accessibility [12]. Particularly during the design stage where many ideas will be tested and discarded, the development process of large range devices could be greatly enhanced with a rapid prototyping system that is flexible enough to automatically manufacture such devices in a cost-effective manner.

The use of polymer as a material for flexure stages can be an alternative method to explore various biological, electrical, and magnetic applications because general polymers are biocompatible and electrically insulating [17–19]. Also, it can provide a large deflection with low force compared to the metal flexures, and the polymeric materials have a higher damping than the conventional metal, which indicates that the polymeric positioning stage fabricated by additive manufacturing (AM) processes acts as an overdamped or critical-damped system and the stage can be driven up to resonant frequency. This can be considered an advantage over the metallic flexure design in applications requiring an insulating material, electromagnetic inertness, biocompatibility, and high damping. In order to become commercially viable, the majority of nanopositioning systems must eventually become compatible with inexpensive, fast, mass-manufacturing processes. Polymer flexures are low-cost, easy-to-fabricate, frictionless, and robust. Furthermore, they are an easy way to incorporate compliance. However, challenges associated with fabricating structures, a number of properties of polymers, such as multi-axial compliance, cross-coupling, thermal insulation, and viscoelasticity/creep have not been studied.

In this chapter, a new displacement sensor based on edge diffraction is first introduced as an alternative position sensor for nanopositioning systems in a simple configuration. Advantageously, the proposed sensing configuration is compact and economical and can be directly embedded into the nanopositioning systems as positioning feedback sensors. The sensing principle and test results have been discussed. In addition, challenges associated with fabricating structures, a number of properties of polymers, such as compliance, out-of-plane motion, and thermal insulation, have been discussed in this chapter.

2   Optical Knife-Edge Diffraction

An electromagnetic waveform diffracts or bends when it strikes the sharp edge of
an obstacles transverse to its propagation direction. The portion of the signal that is
not cut off by the sharp edge continuously travels, but the edge of the signal bends
around the sharp edge [21–26]. The knife-edge techniques are commonly used for
radio communications or optical beam profiling of a laser beam [23, 24]. That effect
is explained by Huygens–Fresnel principle. It states that each point on a wavefront
acts as a new source with secondary wavelets and combination of these secondary
wavelets produces the new wavefront along the propagation direction. That is, it
propagates into the geometric line-of-sight shadow region of the obstacle.

Assuming that the total field above the optical knife edge (OKE) is the incident
field and the light source transverses along the z axis as seen in Fig. 22.2, the
diffracted light to produce results for the edge diffraction can be mathematically
derived. The height h is the displacement from the center of the incident beam to the
OKE along the y_E axis, and L_1 and L_2 are the distances from the laser source to the
OKE and from the OKE to the detector, respectively. The height h can be defined
as a function of its mean height h and its deviation part Δh along the x direction. A
Gaussian beam with a beam diameter α is assumed to be incident on the OKE from
a light source placed a distance L_1 from the OKE. The incident field (E) at $z = 0$ can
be expressed in a spherical wave form as [16],

$$\overrightarrow{E}_0(x, y, z = 0) = E_0 e^{-jk_0 L_1} e^{-\frac{x_0^2 + y_0^2}{\alpha^2}}. \tag{22.1}$$
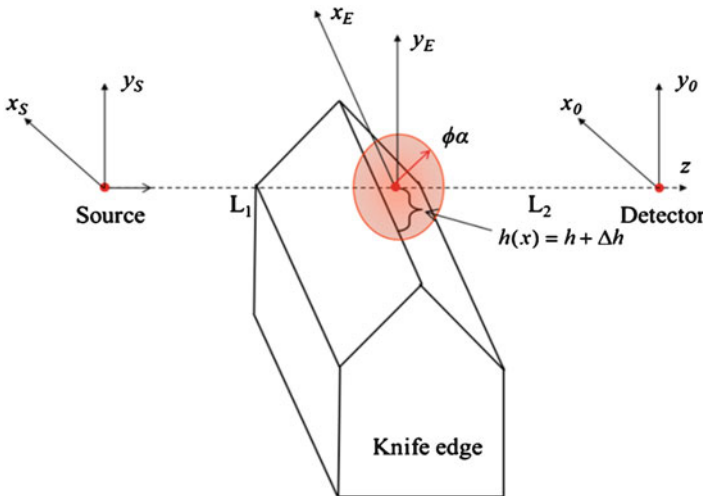
Fig. 22.2 Schematic illustration of OKE diffraction

The first and second terms are an expression of the traveling wave along the propagation direction and the incident spherical wave, respectively. The last term is for an aperture of the field due to the OKE. The transverse wave can be derived by using Fourier transform as Eq. (22.2) and the total field defined as the sum of diffracted and transverse field at a detector position $\overrightarrow{E_t}$ at $z = Z_0$, can be obtained by applying inverse Fourier transform of the transverse wave from the edge as Eq. (22.3) [21].

$$F\left(k_x, k_y\right) = \iint_{\text{Aperture}} \overrightarrow{E_0}\left(x, y, z = 0\right) e^{-j2\pi\left(k_x x + k_y y\right)} dx dy. \qquad (22.2)$$

$$\overrightarrow{E_t}\left(x_0, y_0, z = Z_0\right) = \frac{1}{4\pi^2} \iint_{k-\text{space}} F\left(k_x, k_y\right) e^{j\overrightarrow{k}\cdot\overrightarrow{r_0}} dk_x dk_y. \qquad (22.3)$$

where, the total electric field, the wavenumber, and the distance vector from the incident to the diffracted field $\overrightarrow{E_t} = E_x\overrightarrow{x} + E_y\overrightarrow{y} + E_z\overrightarrow{z}$, $\overrightarrow{k} = k_x\widehat{x} + k_y\widehat{y} + k_z\widehat{z}$, $k_o^2 = k_x^2 + k_y^2 + k_z^2$, $\overrightarrow{r} = x_o\widehat{x} + y_o\widehat{y} + z_o\widehat{z}$ are expressed, respectively.

$$\overrightarrow{E}\left(0, 0, z_0\right) = \frac{jk_0 e^{-jk_0(L_1+L_2)}}{2\pi L_1 L_2}\left(E_x\overrightarrow{x} + E_y\overrightarrow{y}\right)\int_{-\infty}^{\infty}\int_{-h}^{\infty} e^{-a(x^2+y^2)} dy dx$$
$$= \frac{jk_0 e^{-jk_0(L_1+L_2)}}{4\pi L_1 L_2}\left(E_x\overrightarrow{x} + E_y\overrightarrow{y}\right)\text{erfc}\left(-h\sqrt{a}\right), \quad a = \frac{1}{\alpha^2} + jk_0\left(\frac{1}{2L_1} + \frac{1}{2L_2}\right). \qquad (22.4)$$

Assuming that the detector size is sufficiently small that the field of interest will be measured near the z axis then the field measurements are assumed to be taken on the axis ($x_o = 0$, $y_o = 0$, $L_2 > 0$). Therefore, the diffracted field to be measured along the z axis can be defined from the inverse Fourier transform relation of the incident field and written in an error function form as Eq. (22.4) [26]. Where the erfc is the complementary error function of a complex argument. Thus, the total power P induced by the OKE with smooth and sharp surface topography can be calculated by multiplying the diffracted field and the conjugated diffracted field as,

$$P = \overrightarrow{E}\left(0, 0, z_0\right) \times \overrightarrow{E}\left(0, 0, z_0\right). \qquad (22.5)$$

Assuming that the OKE is placed at the center between the laser and the detector for simple calculation, and its surface topography is smooth and sharp. The effects of the wavelength of the light, the distances L_1 and L_2 and the beam width of the laser on the OKE diffraction with respect to the height (displacement) were investigated as seen in Fig. 22.3. The results indicated that the effect of the OKE diffraction becomes more sensitive as the wavelength and distance L_1 and L_2 are shorter, and the beam width rarely affects the result. Thus, it can be found that the short wavelength and the compact measurement setup provide a high sensitivity to the OKE diffraction.
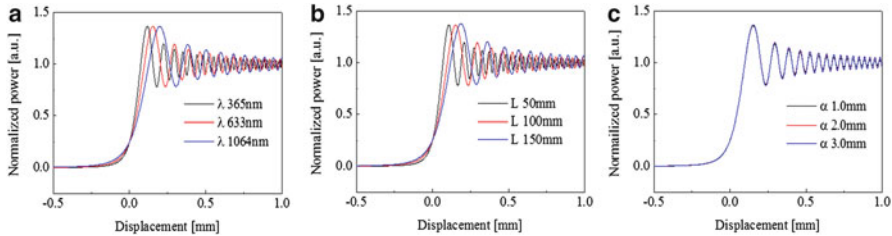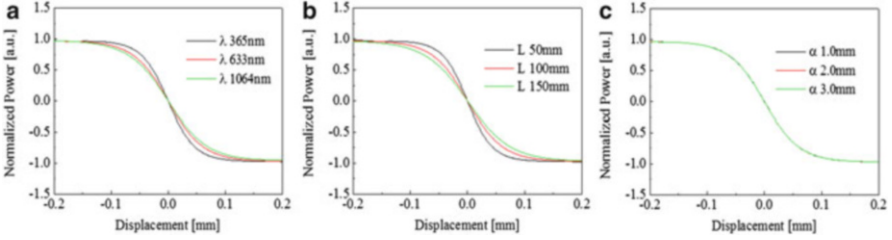
Fig. 22.3 The normalized power curves of the total field calculated with respect to the displacement (h) under given parameter conditions: $L = L_1 = L_2 = 100$ mm and $\alpha = 2.0$ mm (**a**), $\lambda = 633$ nm and $\alpha = 2.0$ mm (**b**) $\lambda = 633$ nm and $L = L_1 = L_2 = 100$ mm (**c**)
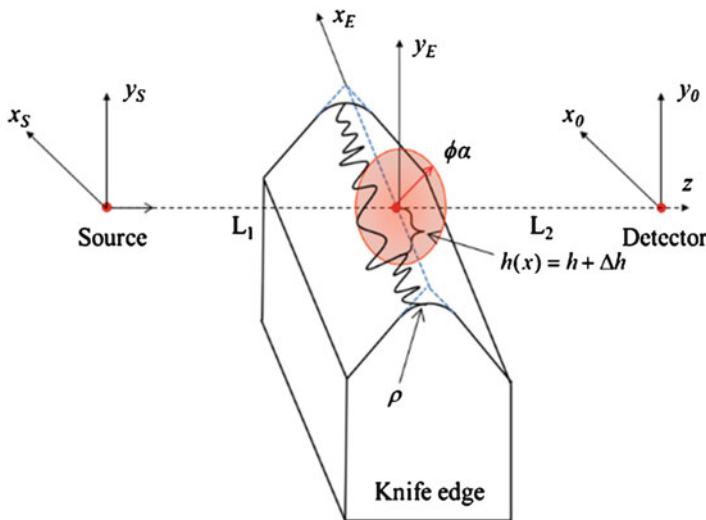

Fig. 22.4 Schematic illustration of OKE diffraction with a rough and dull OKE surface topography: ρ indicates a radius of curvature of the knife edge

Optical metrology is the science and technology concerning measurements with the light. Such measurements can either target properties of light itself or other properties such as displacement, distance, and/or frequency. Optical metrology systems can be extremely precise and are fundamentally limited by surface quality of the optical components, including roughness and form error [2–4, 16]. As referred to in Collin's approach to the rough knife-edge boundary problem [23–26], there has always been an uncertainty relating the effects of edge surface roughness on the diffraction characteristics even though the edge diffraction around a rough knife-edge boundary is modeled. In this study, an approximated model to a rough knife-edge surface was presented. The phase difference between the transmitted and diffracted fields was related to the roughness. It was assumed that the roughness is randomly distributed in nature as seen in Fig. 22.4. The roughness on the knife-edge boundary is assumed to comprise a nonzero mean with Gaussian probability

density function (pdf) and a Gaussian spectrum. Based on this assumption, knife-edge surface roughness is modeled in a pdf form as [27, 28],

$$\text{pdf}\,(\Delta h) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\Delta h^2}{2\sigma^2}} \tag{22.6}$$

where, σ is the standard deviation of Δh. Here, the roughness only along the x axis is considered and independent of the y axis. By substituting Eq. (22.4) into Eq. (22.3), the knife-edge roughness-induced total field $\overrightarrow{E}_R$ can be calculated by [26],

$$
\begin{aligned}
\overrightarrow{E}_R\,(0,0,z_0) &= \int_{-\infty}^{\infty} \frac{jk_0 e^{-jk_0(L_1+L_2)}}{2\pi L_1 L_2} \left(E_x \overrightarrow{x} + E_y \overrightarrow{y}\right) \\
&\quad \int_{-\infty}^{\infty}\int_{-h_C}^{\infty} e^{-a\left(x^2+[h-\Delta h]^2\right)} \text{pfd}\,(\Delta h)\, dh dx d\Delta h \\
&= \frac{jk_0 e^{-jk_0(L_1+L_2)}}{2\pi L_1 L_2} \left(E_x \overrightarrow{x} + E_y \overrightarrow{y}\right) \sqrt{\frac{\pi}{a(2\sigma^2 a+1)}} \int_{-h_C}^{\infty} e^{-a\left(x^2+[h-\Delta h]^2\right)} \text{pfd}\,(\Delta h)\, dh \\
&= \frac{jk_0 e^{-jk_0(L_1+L_2)}}{4\pi L_1 L_2} \left(E_x \overrightarrow{x} + E_y \overrightarrow{y}\right) \text{erfc}\left(\sqrt{\frac{\pi}{a(2\sigma^2 a+1)}}\right).
\end{aligned}
\tag{22.7}
$$

Similarly, the knife-edge roughness-induced total power, P_R, over a given detector size can be calculated by,

$$P_R = \overrightarrow{E}_R\,(0,0,z_0) \times \overrightarrow{E}_R(0,0,z_0)^*. \tag{22.8}$$

In Eq. (22.6), the total field becomes equivalent to that of Eq. (22.4) if the edge topography becomes negligible ($\sigma \to 0$). In the same way, the total power induced by the knife edge with rough and dull edge topography can be calculated by Eq. (22.7). As seen in Fig. 22.5a, the fringe of the oscillation amplitude of the
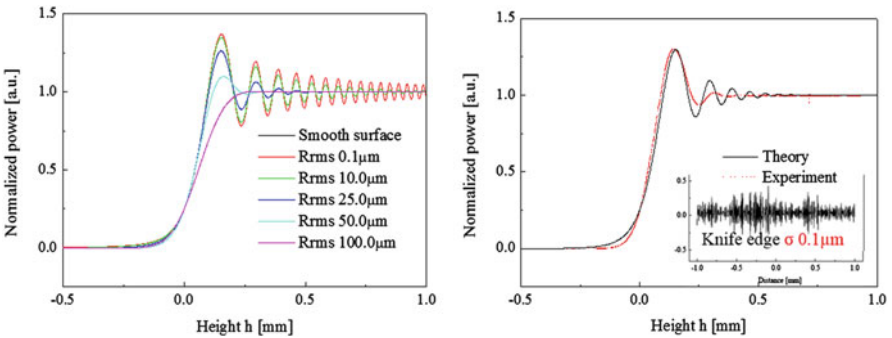
Fig. 22.5 Edge roughness effect of edge diffraction: (**a**) normalized power distributions based on theory and (**b**) experimental result

interferogram reduces as the edge surface roughness increases. The edge diffraction pattern from the commercially available sharp edge with a roughness of 0.1 μm was measured, which shows a similar result with the theoretical estimation as shown in Fig. 22.5b.

3 Novel Displacement Measuring Metrology Based on the OKE

The OKE sensor includes the laser source, beam splitter (BS), reflector, double edge knife edge, and two photodiodes (PD) as seen in Fig. 22.6. The BS-to-knife edge and knife edge-to-PD distances were set to L_1 and L_2, respectively. The light source is separated by a 50:50 at BS. The one light is incident on an upper knife-edge facet through a reflector and another light is incident on lower knife-edge facet. Each light produces interferogram at each PD by a superposition of transverse and diffracted lights from the knife edge, respectively. Therefore, the peak intensity at the first fringe of the interferogram reaches more than 40 % compared to the applied total light intensity [16, 18], provided that the beam diameter is sufficiently larger than the PD size, which indicates that the sensitivity against the motion of knife edge can be increased without any physical amplification due to the knife-edge diffraction principle.

This photovoltaic signals at each PD are then differentially amplified and, consequently, the sensor output $V_{out}(Y)$ along the traveling direction can be detected simultaneously from the combination of the normalized total fields, P_A and P_B.

Where normalized total fields are the intensity detected at PD_A and PD_B as to the y axis displacement, respectively. It uses a difference of P_A and P_B divided by a sum of P_A and P_B in Eq. (22.8),
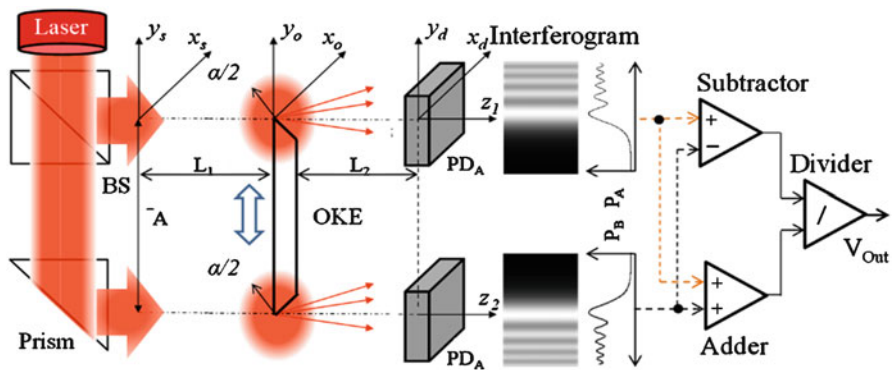
Fig. 22.6 Concept of the OKE sensing system

Fig. 22.7 The normalized sensor outputs with respect to the displacement under given parameter conditions: $L = L_1 = L_2 = 100$ mm and $\alpha = 2.0$ mm (**a**), $\lambda = 633$ nm and $\alpha = 2.0$ mm (**b**) $\lambda = 633$ nm and $L = L_1 = L_2 = 100$ mm (**c**)

$$V_{\text{out}}(Y) = \frac{P_B(Y) - P_A(Y)}{P_B(Y) + P_A(Y)}. \tag{22.9}$$

This method can effectively eliminate the dependency of light intensity and manufacturing tolerance of the knife edge. This measurement method thus can achieve the high sensitivity of the sensor because of the use of the interferogram and differential amplification. The effects of the wavelength of the light, the distances L_1 and L_2 and the beam width of the laser on the OKE sensor with respect to the height (displacement) were investigated as seen in Fig. 22.7. The results indicated that the effect of the OKE diffraction becomes more sensitive as the wavelength and distance L_1 and L_2 are shorter, and the beam width rarely affects to the result. This approach will be used to optimize these parameters for system robustness.

There always exists uncertainty associated with the interference information resulting from the edge diffraction due to the OKE surface topography (roughness, sharpness, shape) and its material properties (dielectric, conducting) as depicted in Fig. 22.4 [16]. The height h is now a function of position along the edge, a function of x and relates the displacement along the z direction. As the OKE becomes rough and dull, the edge diffracted field becomes more incoherent and the interference becomes weaker. It is thus very crucial to investigate the effects of the topography on the OKE diffraction of an incident wave in a quantitative way. However, the researches regarding quantitative characterization of the roughness and sharpness of the OKE have not been well-documented.

Our research group firstly introduced a novel displacement sensing technique utilizing the OKE diffraction as an alternative position sensor for nanopositioning systems in a simple configuration as illustrated in Fig. 22.8. The proposed measurement system utilizes the interference superposed by two waves, transverse light and the diffracted light from the OKE and a differential amplification signal processing technique in order to eliminate the potentially much greater noise signals from power sources and environmental effects [2, 3]. The measurement system consists of laser source, two knife edges, two PD and axillary optics components is constructed on a nanopositioning stage in a simple configuration. The laser light is split 50/50 by BS and the transmitted and reflected lights are incident on each
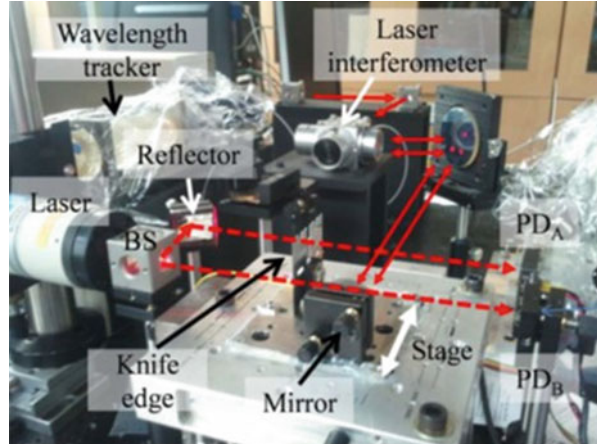
OKE mounted on the stage parallel to the traveling direction. While the stage is in motion, the direct transverse and diffracted lights at two OKEs are superposed and introduce interference, which is measured by two photodetectors (PD_A and PD_B) in a differential configuration. Consequently, the displacement along the traveling direction can be detected simultaneously by using PD_A and PD_B. The methods for displacement measurement (ΔD) can be categorized into two types as,

$$\Delta D = \frac{1}{S}\frac{P_A - P_B}{P_A + P_B} \tag{22.10}$$

where S is a sensitivity of the sensor. P_A and P_B are the total power measured at PD_A and PD_B, respectively. It uses a difference of P_A and P_B divided by a sum of P_A and P_B as expressed in Eq. (22.9), which can effectively eliminate the dependency of light intensity. This measurement method is considered to achieve the high resolution and sensitivity of the sensor because the interference is highly sensitive to the displacement of the OKE. Moreover, it can enhance the sensitivity and eliminate crosstalk of the sensor because it provides a gap-free application that a position sensitive detector (PSD)-based measurement system cannot achieve [2, 3, 29].

Our research group preliminarily tested the performance of the proposed the OKE displacement sensor by implementing the sensor with a lab-built flexure stage. A stabilized He–Ne laser (λ 633 nm) as a coherent light source was used. The laser beam is separated 50/50 at the BS. A transverse beam and reflected beam are incident onto the OKE fixed on the stage, respectively, and a small fraction of light is diffracted at the OKE. A transverse beam and diffracted beam experience the interference due to phase matching and mismatching, and those signals are measured at both PD_A and PD_B in a differential amplifier configuration and collected with a LabView data acquisition board under the ambient environment condition. The distance L_1 between the laser and the OKE and the distance L_2 between the

Fig. 22.9 Calibration results
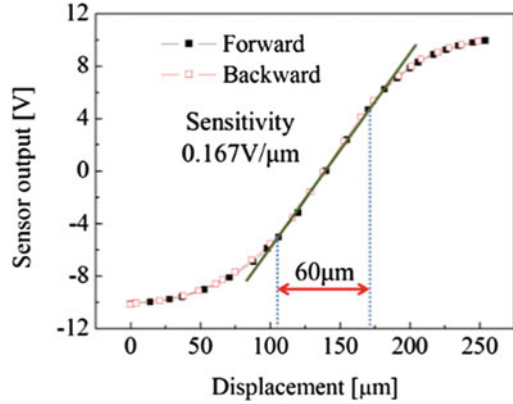
OKE and PDs are set to 100 mm, equally, and the beam width is 2.0 mm. The laser interferometer and a one-dimensional PSD were used to measure the displacement of the stage and compare these results with that of the proposed OKE displacement sensor.

The proposed sensor was calibrated according to forward and backward directions and it showed a high linearity within approximately 60 µm range and a high sensitivity of 0.167 V/µm (Fig. 22.9). In this experiment, a laser beam diameter is limited to ∼0.5 mm, so that the measuring range and sensor sensitivity were determined by the laser beam diameter. A look-up table can be made to increase the measuring range, and larger the beam diameter is, the larger measuring range is achieved. Here the linear range and the sensitivity were chosen based on the stroke of the precision stage used in the experiment. It is possible that the measurable range of the sensor becomes shorter or longer by increasing or decreasing its sensitivity according to the stroke of the stage.

The displacement information was measured from three displacement sensors, laser interferometer, PSD, and OKE sensor, at the same time while the stage was in 1 Hz sinusoidal motion under the stroke conditions: 25 and 0.03 µm. As seen in Fig. 22.10, three sensors showed similar results a stroke of 25 µm. For a stroke of 0.03 µm, the PSD signal showed fluctuation and a higher noise-to-signal ratio whereas the edge diffraction sensor showed similar results with that of the laser interferometer with respect to all conditions. In addition, the signals of three sensors were measured for 30 s. to test the stability of each sensor while the stage was held stationary as seen in Fig. 22.11. It can be seen that the noise level of PSD is approximately 200 nm. While, the noise level of the proposed OKE diffraction sensor was near 5.2 nm. The output of the laser interferometer showed the fluctuation because the wavelength tracker was disabled in order to make the experiment condition of each sensor equal. As a result, the edge diffraction sensor showed a better stability than the other two sensors, and it can be concluded that the proposed edge diffraction sensor is comparable to the laser interferometer.
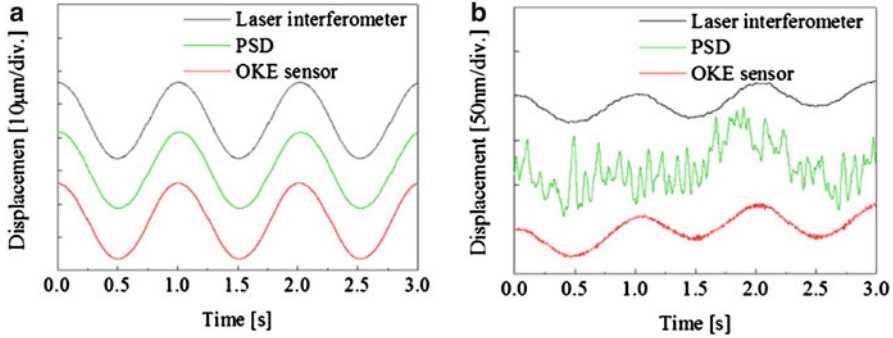
Fig. 22.10 Measurement results under stroke of 25 μm (**a**), 6 μm (**b**), 0.3 μm, (**c**) and 0.03 μm (**d**) conditions and comparison to the measurement results of the laser interferometer and PSD [16]

Fig. 22.11 Stability test: Three instruments (LI, PSD, and OKE sensor) were tested under stationary condition (power off) [16]
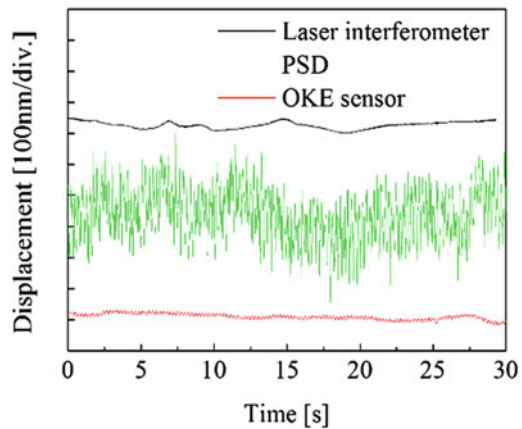
4   Additive Manufactured Compliant Mechanism

For precision translation and positioning systems, most of the mechanisms manufactured today exploit the versatile capabilities of CNC machine tools and EDMs. Nearly all flexure systems are fabricated using conductive materials, primarily metal. Flexure system fabrication uses methods planar processing methods to remove material. However, metal-based positioning stages are not only time consuming and expensive to manufacture but also difficult to use in chemical, electrical, and electromagnetic environments, and in biotissue-handling work. Metallic flexure systems are subject to corrosion, interference with electromagnetic field distributions and can suffer from biocompatibility issues [14, 17, 19]. In addition, system bandwidth of the commercially-available metal-based flexure stages is approximately 100 Hz because most flexure stages have a low first-resonance frequency (less than 400 Hz) [30, 31]. Recently, with progress of microelectromechanical system (MEMS) technology, many MEMS motion stages have been introduced [32,

33]. However, the total displacement is limited to a few tens of μm, even still, they are promising for high bandwidth and high resolution. The use of polymer as a material for flexure stages can be an alternative method to explore various biological, electrical and magnetic applications because generally polymers are biocompatible and electrically insulating. Also, it can provide a large deflection with low force compared to the metal flexures, and the polymeric materials have a higher damping than the conventional metal, which indicates that the polymeric positioning stage acts as an overdamped or critical-damped system and the stage can be driven up to resonant frequency.

The use of polymer as a material for flexure stages can be an alternative method to explore various biological, electrical and magnetic applications because general polymers are biocompatible and electrically insulating. Few studies on polymeric materials and those fabrication methods for positioning stage applications have been introduced to solve troublesome issues raised in conventional technologies [2]. In addition, the polymeric materials have a higher damping than the conventional metal, which indicates that the polymeric positioning stage acts as an overdamped system and that the stage can be driven up to resonant frequency. Thus, the polymeric flexure can be considered as being advantageous to the metallic flexure design in applications requiring an insulating material, electromagnetic inertness, biocompatibility, and high damping. A few polymeric flexure stages have been proposed. Chae [19] reported a transparent nanopositioner made of polymer (Topas 5013 COC) by injection molding and Hoover [20] introduced a low-cost two-axis displacement compliant mechanism made of polydimethylsiloxane (PDMS) by molding. Lee tested the performance of a 3D printed compliant mechanism with the integrated displacement sensor [17, 18].

This chapter addresses the use of additive manufacturing to create polymeric precision stages for the purpose of a large travel range and low cost. Additive manufacturing provides three dimensional and free from manufacturing which promises to be another major contributor to further innovation in this field. Since the 1980s the development of additive manufacturing technologies has been driven by companies looking to produce parts rapidly without the need of dedicated tooling. It provides design engineers with a new tool during the product development cycle and allows for shorter product lead times [33–35].

In this chapter, the concept of a polymeric flexure mechanism for a long working range and ultralow cost by using AM technology has been tested. The double compound notch flexure mechanism made of photopolymer is explored to demonstrate this concept and the effects of its geometry on the flexural deflection and stress are investigated. The performance and limitation of monolithic elliptical hinges is analyzed and designed based on a single rotational degree-of-freedom mechanism.

A double compound notch type flexure mechanism (DCNFM) was used in the research. This double compound kinematic design is well-known in precision engineering [12, 17, 18]. It has a symmetric design configuration consisting of four compound springs attached at the shuttle (moving platform). The use of four compound springs joined together constrains the stage motion to prevent deviation
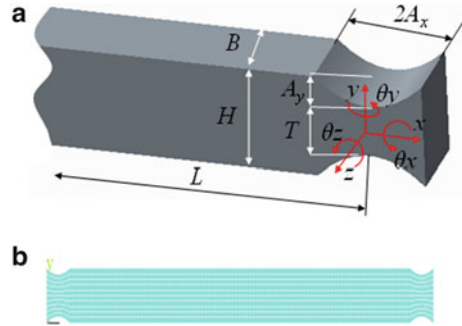
from rectilinear motion. The translation of the shuttle is controlled by the elastic
deformation of two identical four-spring mechanisms. The use of two symmetrical
four-spring mechanisms offers an additional advantage in that the system can self-
compensate for thermal expansion [12].

Each spring was designed with elliptical flexure hinges at both ends because
these hinges are commonly used in compliant micromotion stages which required
high precision of motions [37–41]. This type of flexure hinges is precise in rotation
because their center of rotation does not displace as much as other hinges such
as the leaf-type [38] and the corner-fillet [39]. In addition, the elliptical flexure
hinge for the flexure stages can avoid abrupt changes in the cross-sectional flow area
while generating typically pure rotation. Considering the elliptical hinge structure
in Fig. 22.12, the height y with respect to a position x on the stage can be expressed
by

$$y(x) = A_y + \frac{T}{2} - A_y\sqrt{1 - \frac{x^2}{A_x^2}}. \tag{22.11}$$

where A_x and A_y are the length of the hinge, T is the thickness of the hinge, B the
width of the hinge, L the length of the flexure, and H is the height of the flexure.
Here the performance and limitation of monolithic elastic hinges was analyzed and
designed based on a single rotational degree-of-freedom mechanism. Following the
analysis of Paros and Weisbord [40], the approximated displacement of the elliptical
hinge may be expressed by integrating the bending equation as

$$\theta_z(x) = \int_{-A_x}^{A_x} \frac{M_z}{EI_z}dx = \int_{-A_x}^{A_x} \frac{12M_z}{8EB[y(x)]^3}dx. \tag{22.12}$$

where I_z is the second moment of area, M_z the bending moment, E the elastic
modulus, θ_z the angular deflection about the neural axis. Substituting Eq. (22.10)
into Eq. (22.11) and $x = A_x \sin\theta$, Eq. (22.11) can be rewritten in a compliance $C_{\theta_z M_z}$
form by
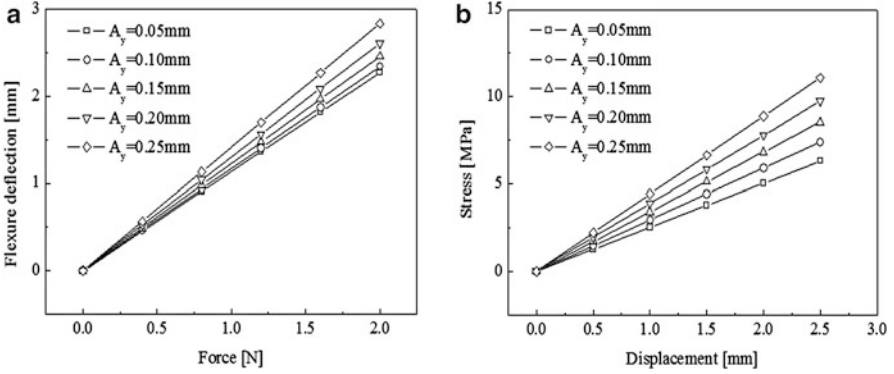
Fig. 22.13 FEM analysis result of flexure: (**a**) flexure deflection curves according to various A_y conditions under the applying forces and (**b**) maximum stress (at the hinge) curves according to various A_y conditions under the displacement condition. The A_x is fixed 0.75 mm

$$C_{\theta_z M_z} = \frac{\theta_z}{M_z} = \frac{3A_x}{2EBA_y^3} \int_{-\pi/2}^{\pi/2} \frac{\cos\theta}{\left[1 + \frac{T}{2A_y} - \cos\theta\right]^3} d\theta. \qquad (22.13)$$

The compliance of the elliptical hinge was estimated with respect to the ratio $\varepsilon = A_x/A_y$ varying from 1 to 10 with different A_y conditions as seen in Fig. 22.13a. It was found that the compliance of the elliptical hinge is linearly proportional to the increase in the ratio under the same A_y condition, and the thicker A_y is, the higher the compliance is found.

Flexure hinges experience both bending and shear stress and their magnitude depends on the geometry of the flexure mechanism and the nature of the applied loads. The stress at the notch hinge is calculated from a nominal stress by Ling [41] and Peterson [42] in terms of a stress concentration factor K_t, applied bending moment M_z and the true stress σ_z as Eq. (22.13). The geometry of the hinge is considered as the circular notch type for simplicity. The true stress at the thinnest area of the hinge is estimated from the bending moment for a given displacement and substituting Eq. (22.11) into Eq. (22.13) because the applied bending moment is unknown [12].

$$\sigma_z = K_t \frac{6M_z}{T^2 B} = \frac{4EA_x^2}{f(\beta)T^2}(1+\beta)^{9/20}\theta_z,$$
$$K_t = \left(1 + \frac{T}{2A_x}\right)^{9/20} = (1+\beta)^{9/20}, \qquad (22.14)$$
$$f(\beta) = \frac{1}{2\beta+\beta^2}\left(\frac{3+4\beta+2\beta^2}{(1+\beta)(2\beta+\beta^2)} + \frac{6(1+\beta)}{(2\beta+\beta^2)^{3/2}}\tan^{-1}\left(\sqrt{\frac{2+\beta}{\beta}}\right)\right).$$

The maximum stress at the hinge was estimated with respect to a given displacement under different A_y conditions as seen Fig. 22.13b. The results demonstrate that the stress linearly increases as the displacement increases and A_y decreases. In addition, the maximum displacement of the flexure is limited to approximately

0.04 rad under the given conditions due to the flexural strength of the polymer used (which is approximately 62 MPa [43]). Here the maximum displacement 0.04 rad was set as a design constraint for the flexures. In a double compound notch type flexure mechanism, the displacement of each flexure becomes a half of the total displacement of the shuttle under an identical flexure geometry condition [12]. Based on these analysis results, the double compound notch type flexure was designed to produce the total ±2.0 mm linear displacement: $\varepsilon = 5$ ($A_x = 0.75$ mm, $A_y = 0.15$ mm), $T = 1.2$ mm, $H = 1.5$ mm, $B = 15.0$ mm, $L = 25.0$ mm. The compliance and max stress of each designed flexure were estimated 0.33 rad/Nm and 61.9 MPa at a maximum displacement, respectively.

The layer-to-layer printing thickness, printing temperature and curing temperature, and time were set to 125 μm, 29 °C, room temperature and 50 min. with UV light, respectively. According to the datasheet [43], the elastic modulus of photopolymer is 1.85 GPa, but the strain–stress test result showed 795 MPa as seen in Fig. 22.14a. In order to evaluate the fatigue properties of this AM material a new test machine was required. The test machine capable of testing the fatigue characteristics of AM material was built in-house due to the low load and high-deflection characteristics of the AM materials. The design process and requirements were guided closely by ASTM D7774-12, the Standard Test Method for Flexural Fatigue Properties of Plastics [35] along with ASTM D790-10 [36]. These specifications provide good insight into design specifics and standard approaches. The test machine uses three-point bending, chosen because the testing would not exceed 5 % strain, and would be tested at constant deflection (strain-based). AM samples (W 127 mm by H 12.7 mm by T 3.3 mm) were prepared. The test machine allows cyclical sinusoidal (5 Hz) loading of the test sample at constant deflection. Load for each cycle is recorded, when loading of the sample decreases by 10 % of the original load, the part is said to reach failure.

Test samples printed along two different directions were tested. Curve fitting is performed via a linear function in log scale. The samples showed a fatigue life cycle over 1.0E5 at 0.2 % strain condition as seen in Fig. 22.14b. At present, we are
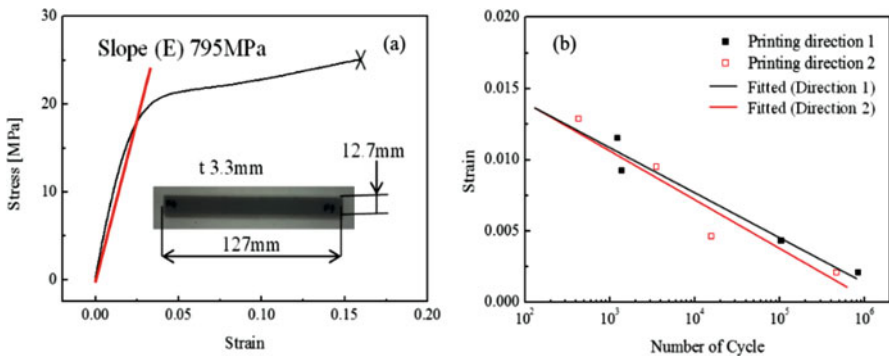
Fig. 22.14 Material property characterization: (**a**) stress–strain test and (**b**) fatigue life cycle test

studying the fatigue characterization of the AM flexures, which will be used as a design principle in AM flexures.

The flexure stage was fabricated by a stereolithography process (Projet® 6000HD, 3D Systems Inc.) as seen in Fig. 22.15. The photosensitive photopolymer (Visijet® SL Tough, $8/in.³) was used. It took less than 1.5 h to fabricate the whole monolithic structure (6.4 in.³). As seen in Fig. 22.15, the flexure hinge was successfully fabricated; $A_x = 0.75$ mm, $A_y = 0.15$ mm, $T = 1.2$ mm, $H = 1.5$ mm.

The surface quality of the AM stage along the height direction was evaluated by using a high resolution optical scanning microscope (VHX-2000, Keyence Inc.) because the surface quality of flexures will introduce an extra stress concentration to the flexure. This affects the fatigue strength significantly due to possible crack nucleation and propagation on the surface [3, 23, 24]. The top surface of the printed flexure was smooth and shiny as shown in Fig. 22.15. The surface profiles along the vertical direction and horizontal direction, were measured as shown in Fig. 22.16.

The AM flexure showed approximately 6 and 1 μm surface waviness along the vertical and horizontal directions, respectively. Taking into consideration of the typical EDM machining tolerance (5–10 μm), it was thought that the AM approach promises to nanopositioning device applications: small tolerance design (6 μm), low printing time (1 h), low-cost and high-design freedom. Also, the pitch of the surface
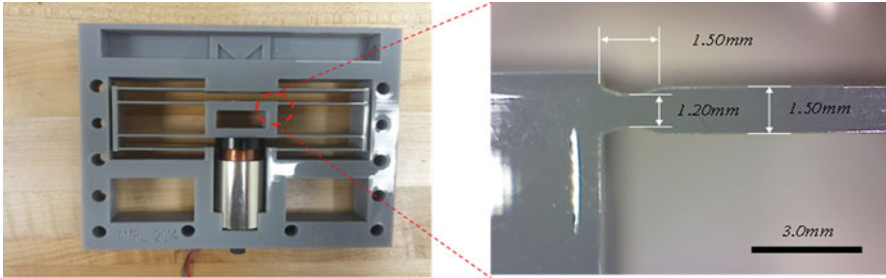
Fig. 22.15 Image of the flexure stage and detailed image of the hinge
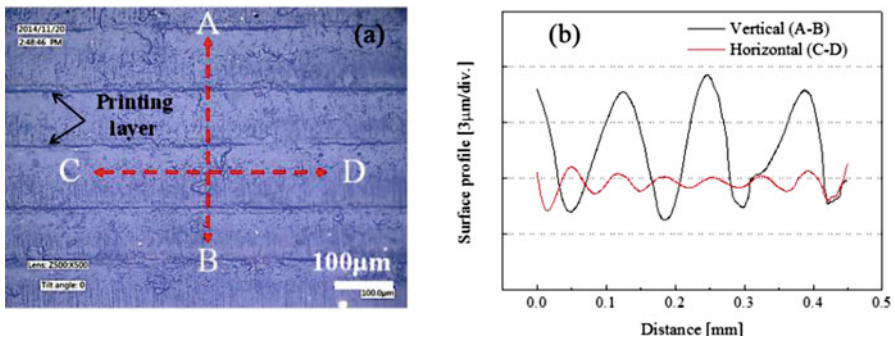
Fig. 22.16 Surface topology of the flexure: 500X microscope image (**a**) and surface profile (**b**)

profile along two directions indicates that the printing layer-to-layer thickness along the vertical direction is approximately 120 μm and the printing condition along the horizontal direction is approximately 70 μm.

It is expected that the study on the AM tolerance and fabrication conditions such as the layer-to-layer printing thickness, printing speed, printing temperature, and curing temperature and time, is in needed for future precision stage applications.

5 Single-Axis Nanopositioning System

The optical knife-edge sensor (OKES) was successfully embedded with DCNFM as shown in Fig. 22.17. With use of the lensed PDs, the measuring range became longer. The CS (10 nm resolution with 500 μm range, Lion Precision Inc.) was used to calibrate the OKES and compare with its performance and control effectiveness of the OKES-based nanopositioning system. The analog circuit was created for a PID controller, motor driver and sensor amplifier. The OKES measuring range was set similar to the CS working range to compare the two sensors at the same condition. As seen in Fig. 22.18, the OKES was calibrated by using the CS and showed a sensitivity of 36.0 mV/μm with high linearity. Positioning control was achieved by PID turning and the control block diagram is shown in Fig. 22.19, where i and i' are the input and output signals, respectively. The offset, C, was set to increase the stiffness of the flexure on purpose and the current feedback loop was made to eliminate crossover distortion that occurs in push–pull amplifier. The total displacement was measured by applying the input voltage under the PID positioning control as seen in Fig. 22.6. This result indicates that the total working range of the stage is approximately 600 μm and the sensitivity of displacement versus input voltage is 29.23 μm/V within a full range.

The 100 μm step response of the nanopositioner was tested as seen in Fig. 22.20a. The CS is represented by CS. The positioning control was achieved with the OKES
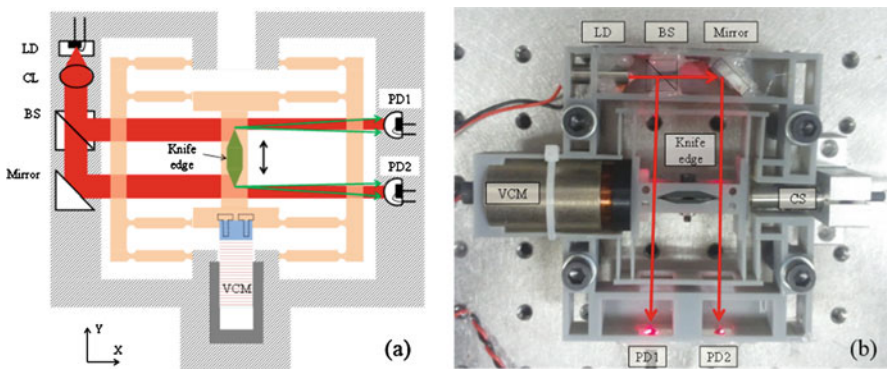
Fig. 22.17 Nanopositioning system integrated with OKES: (**a**) schematic and (**b**) image [18]

Fig. 22.18 Calibration result
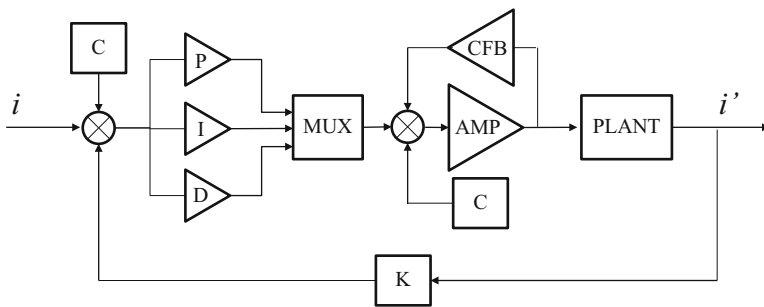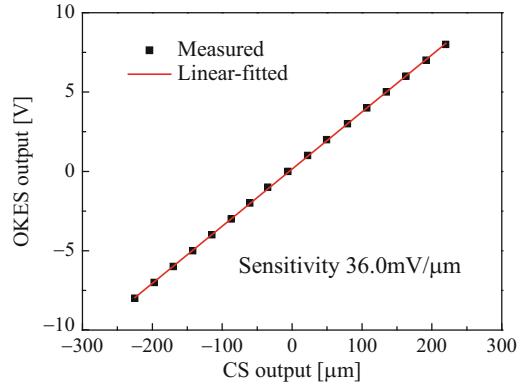
Fig. 22.19 Positioning PID control block diagram: offset (C), Proportional/Integral/Derivative gain (P/I/D), Adder (MUX), VCM amplifier (AMP), current feedback (CFB), OKES sensitivity (K)
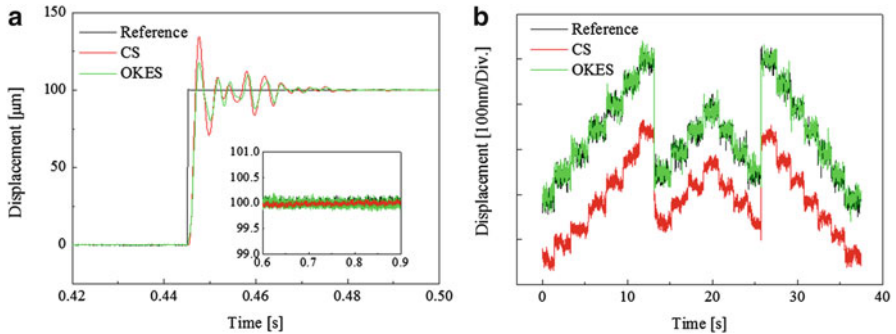
Fig. 22.20 Experimental results: (**a**) 100 μm step response and (**b**) stepwise positioning resolution test

and a CS was used for comparison. The OKES result showed a good agreement with the CS results in terms of the positioning pattern to the target position such as rising time (0.5 ms), overshoot and settling time (30 ms) even though the measurement directions of the two sensors are different. The OKES measures the displacement perpendicular to the driving direction of the stage. On the other hand, the CS
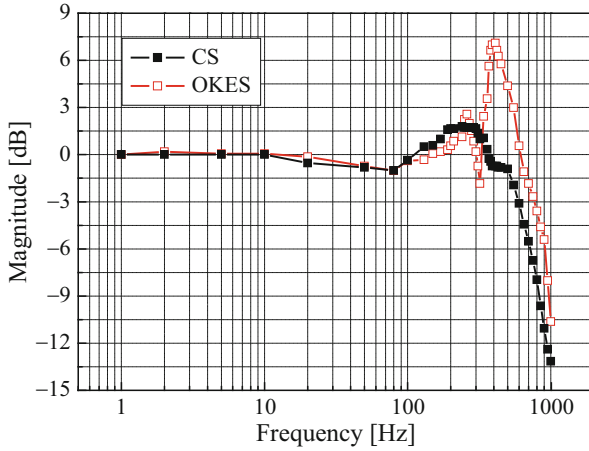
Fig. 22.21 Frequency response curves

Fig. 22.22 Hysteretic
behavior between input
voltage and displacement
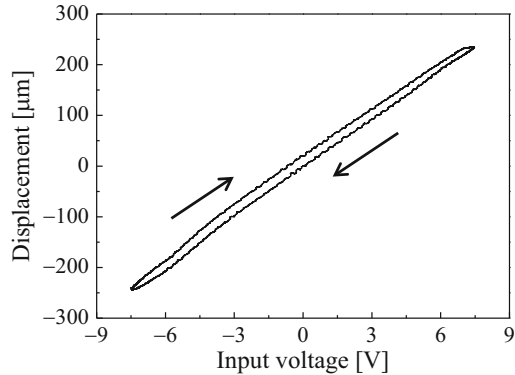
measures along the driving direction. The OKES and input signals showed a perfect overlap in a steady-state region (0.6–0.9 s), and the CS signal appeared within a noise range of the OKES and input signals (Fig. 22.21).

The input voltage-dependent displacement curves of the flexure stage had a nonlinear course that is subjected to hysteresis as shown in Fig. 22.22. It was thought that the error is due to PID gain mismatching as well as the hysteresis and nonlinearity of the VCM. In the previous research [17], AM flexure stage was positioning feedback-controlled by using CF, and the hysteresis characteristics of AM flexure mechanism was discussed. The maximum hysteresis error (difference between loading and unloading conditions) was approximately 20 μm for CS-based control systems. Here maximum hysteresis error was approximately 25 μm. Hysteresis can negatively affect the positioning control loops, decreasing the performances or even leading to instability. Thus, a proper hysteresis model and control algorithm has to be used in feedback control loops to compensate for the nonlinearity. Hysteresis compensation will be studied for the next work.

The stepwise inputs, 50 nm, were given to test the positioning resolution of the stage and CS was placed for a comparison with the OKES as shown in Fig. 22.20b. These results indicate that 50 nm positioning control was successfully achieved. The CS showed a good agreement with the OKES over the full range, but the CS showed small discrepancy in displacement for the forward and backward direction. It was considered that this discrepancy results in sensor nonlinearity, sensor installment error due to Abbe or cosine error, or rotational motion of the stage. As a result, the proposed nanopositioner showed 50 nm positioning resolution with 500 μm working range.

The dynamic characterization of the stage was investigated by obtaining the frequency response curves as seen in Fig. 22.22. Two sensors showed similar results below 100 Hz and a resonance peak of the stage at 280 Hz. However, the OKES showed a large second resonance peak at 400 Hz. On the other hand, the CS showed no significant peak around 400 Hz. It was thought that the first resonance peak is due to the flexural mode of the stage and second resonance peak is related to torsional mode of the stage because the flexure stage typically has a coupled motion with flexural and torsional modes [12, 14, 44, 45]. It was explained that the CS measures the capacitance change of an effective area (ϕ3.2 mm) at a single point and it becomes less-sensitive to the torsional motion if it is installed along the driving axis with no offset distance. On the other hand, the OKES measures two points (22 mm interval) around the center of the stage and it can be sensitive to the torsional motion. As such, angular motion, especially for high frequency signal, has an influence on OKES readout, which does not mean that OKES does not perform well. The OKES can measure the angular motions with high sensitivity which the CS cannot measure.

6 Two-Axis Nanopositioning System

The *XY* nanopositioning system (150×150 mm^2, 20 mm thickness) based on parallel kinematic and monolithic configuration was designed to yield a compact layout as seen in Fig. 22.23 and provide more than ± 1.0 mm range and fabricated by using stereolithography process [17]. From the open-loop testing, the *XY* nanopositioning system was characterized: natural frequency (55.6 Hz/52.6 Hz) and damping ratio (0.14/0.09) along *X* and *Y* axis, respectively, as shown in Fig. 22.24. The two OKES were embedded into the stage and measure *XY* displacement to control the positioning of the stage in real time. The LD (5 mW, λ 650 nm, α 5.0 mm) light is 50:50 separated at BS and travels to each OKES with surface roughness ∼0.1 μm. The VCMs were used as actuators. Four photodetectors (PD, 3×3 mm^2) were installed at the center of the stage to receive the transmitted and diffracted light behind the OKE with the distance *L* 10 mm, and PD electronics was set to a bandwidth of 10 kHz.

The theoretical sensor output calculated in Eq. (22.8) shows a sensitivity of 0.667×10^{-3} μm^{-1} and high linearity (nonlinearity 0.0018 %) within ± 1.0 mm

Fig. 22.23 Schematic illustration of the OKES-embedded nanopositioning system: The *yellow lines* indicate the laser beam path and the *red arrows* indicate the light propagating direction, respectively. The VCMs were aligned along the moving axis of the stage without offset to mitigate Abbe error, and two OKESes were setup perpendicularly to one another. The flexure was designed: 20 mm long, 15 mm wide, and 1 mm thick
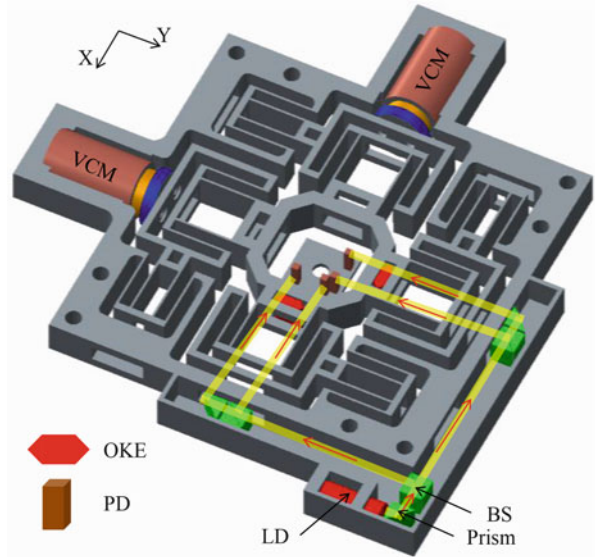
Fig. 22.24 Open-loop step response results
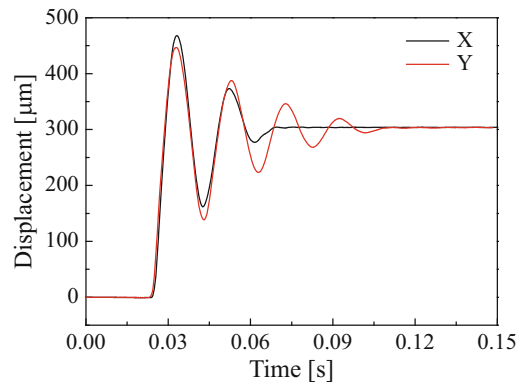
range as seen in Fig. 22.25. The *XY* nanopositioning system was proportional–integral–differential (PID)-controlled with the OKES and CS was used for comparison and calibration. Each OKES showed sensitivity 8.99 mV/μm and 9.02 mV/μm within ±500 μm (CS max. measurement range) with nonlinearity 0.60% and 0.73% along the *X* and *Y* axes, respectively. A 300 μm step test of the *XY* stage was performed as shown in Fig. 22.26. The OKES and CS results showed good agreement and the steady-state positioning root-mean-square (rms) noise, which is also representative of the resolution, was 21.5 nm and 19.3 nm along *X* and *Y* axes, respectively. We think that the discrepancy of *XY* step response results from AM fabrication tolerance (flexure width, thickness and surface roughness), material property non-uniformity or the different VCM characteristics including current amplifiers. As a result, the fundamental sensing limit of the OKES in the
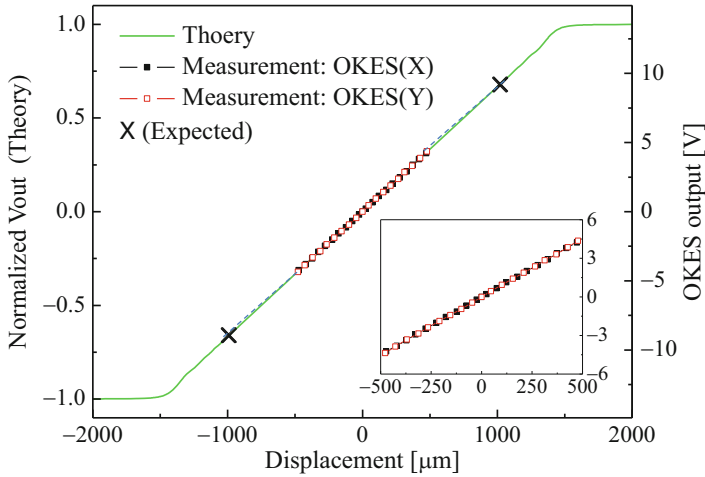
Fig. 22.25 Normalized sensor output calculated with respect to the OKE shift along the y_0 axis. The theoretical estimation result works for both X and Y axes because two OKESes were setup independently to one another along the X and Y axis. To illustrate good agreement between theory and experimental results more clearly, the scale of the OKES output was adjusted
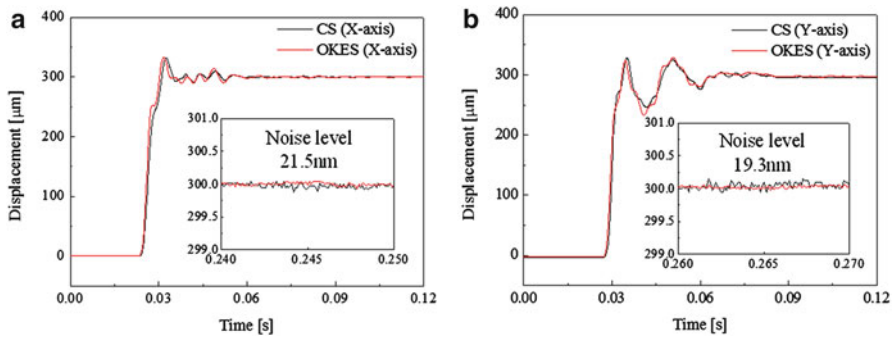
Fig. 22.26 Step response curves of the X (**a**) and Y (**b**) axes

present experiment setup can be determined by the photonic shot noise (N_S) in Eq. (22.14) [46, 47],

$$N_S = \sqrt{\frac{\Delta R}{\Delta f}} \qquad (22.15)$$

where, Δf and ΔR are the measurement bandwidth and resolution, respectively. The photonic shot noise indicates the variation of the sensing signal as a function of its resolution and measurement bandwidth. Thus, the fundamental sensing limit of the OKES was estimated X 0.19 nm/$\sqrt{Hz}$ and Y 0.23 nm/$\sqrt{Hz}$ for ± 1.0 mm working
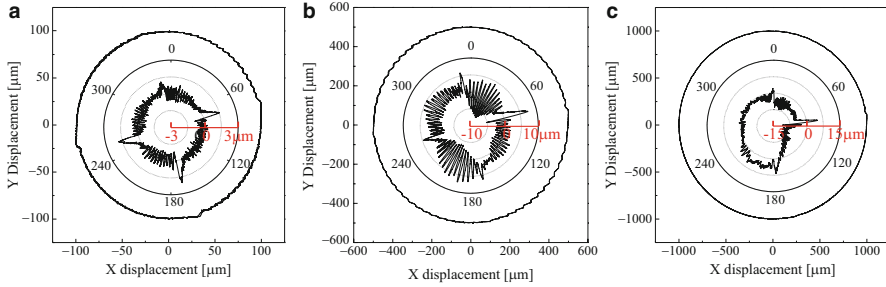
Fig. 22.27 Results of circular motion tracking at 1 Hz: diameter 0.2 mm (**a**), 1.0 mm (**b**) and 2.0 mm (**c**). The trajectory errors were plotted in a polar coordinate

range. This noise can be reduced by using a sufficiently bright light source with a large beam diameter relative to PD [2, 3].

The nanopositioning system was PID feedback-controlled to draw a 0.2/1.0/ 2.0 mm diameter circle at 1 Hz while sending two sinusoidal reference signals with a magnitude of 0.1/0.5/1.0 mm and relative phase difference of 90° along the both axis. The measured rms trajectory errors were approximately 0.77/2.50/3.50 μm for each circular motion and these tracking errors correspond to 0.38/0.25/0.17 % over each range, respectively (See Fig. 22.27). We think the following are responsible for the trajectory errors: (1) magnitude and phase errors in the actual trajectory along each axis with respect to the sine and cosine inputs, (2) higher order harmonics of the given frequency, (3) electronics noise, and (4) control limits due to PID gain mismatching and VCM performance. Once displacement sensors are installed on the stages, the sensing system accuracy is determined not only by the sensor but also by the stage and controller such as hardware the dynamic characteristics, control noise, and/or electronics noise. Thus, it was considered that the proposed OKES resolution must be less than 20 nm.

7   Long-Term Stability Test

A preliminary investigation of thermal effects [48] was performed and the temperature was measured at the following three points: at the VCM, at the frame near VCM and at the frame far from VCM (PSD area). The temperature was monitored while the flexure was being subjected to cyclic loading of 3.6 million cycles in a temperature-controlled lab environment. The temperature of VCM and the frame near VCM rapidly increased and saturated at 6.0 and 2.5 °C after approximately two million cycles as shown in Fig. 22.28. However, the temperature at the frame far from VCM remained constant with a small fluctuation. The results of step response before and after 3.6 million cycles were compared, and positioning accuracy and

Fig. 22.28 Temperature measurement of the stage during 3.6 million cycles
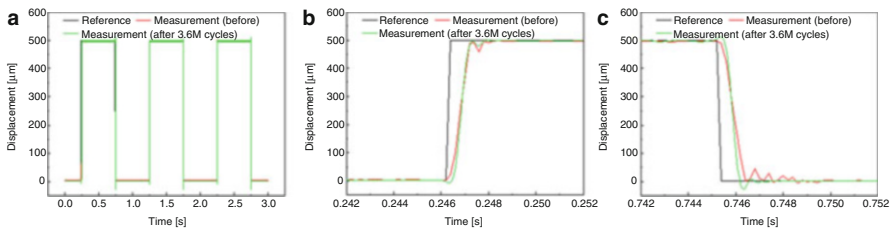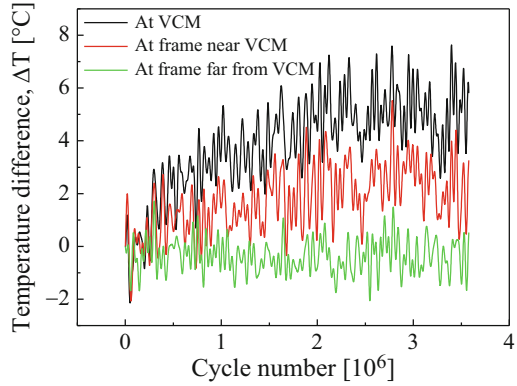
Fig. 22.29 Results of step response of the stage before and after 3.6 million cycles: full scale (**a**), forward direction (**b**) and backward direction (**c**)

control performance showed a similar result as seen in Fig. 22.29. The temperature change and material property change of the flexure will be quantitatively addressed, and its fatigue characteristics will be studied in future.

8 Conclusions

This book chapter presented a design principle of the OKES by deriving a mathematical model and characterized the OKES performance in terms of working range, positioning accuracy, resolution, linearity, bandwidth and control effectiveness with the nanopositioning systems. From the theoretical approach, it was found that this sensing technique is advantageous to achieve high linearity over a millimeter-level range and high bandwidth. The bandwidth of the sensor is limited by electronics bandwidth because the OKES includes PDs, but it can easily exceed MHz levels. The results showed that the OKES is capable of a large range and nanometric resolution for single-axis or multi-axis operation. It offers compact size and low cost, as such this sensor can be a good alternative to the CS and optical encoders or even better sensor selection in nanopositioning applications such as high resolution microscopy, surface profilers or coordinate measuring machines. Also, the polymer-

based monolithic double compound notch type flexure stage was designed by investigating the effect of its geometry on the compliance and the stress of the flexure and successfully fabricated by using additive manufacturing process at an ultralow cost. For the future work, it will be investigated on the fundamental limits of the OKES by characterizing the effects of linearity and sensitivity on the edge diffraction of an incident wave due to the OKE topography, and it is expected that the use of a high resolution digital controller can contribute to a further improvement in positioning resolution and accuracy and a proper hysteresis model and control algorithm development can be a bridge to achieve high positioning resolution. Also, the mechanical property and fatigue characteristics of the AM stages according to the temperature and material property irregularity will be quantitatively addressed.

References

1. Slocum AH (1992) Precision machine design. Prentice-Hall, Englewood Cliffs
2. Lee CB, Lee S-K (2013) Multi-degree-of-freedom motion error measurement in an ultraprecision machine using laser encoder—review. J Mech Sci Technol 27(1):141–152
3. Gao W (2010) Precision nanometrology: sensors and measuring systems for nanomanufacturing. Springer, London
4. Webster JG (1999) The measurement, instrumentation, and sensors handbook. CRC Press LLC, Boca Raton
5. Riemer O (2011) Advances in ultraprecision manufacturing. Proc Jpn Soc Precis Eng
6. Lee CB, Kim G-H, Lee S-K (2011) Design and construction of a single unit multi-function optical encoder for a six-degree-of-freedom motion error measurement in an ultraprecision linear stage. Meas Sci Technol 22:105901
7. Lee CB, Kim G-H, Lee D-J, Lee S-K (2011) A six-degree-of-freedom motion characterization of ballscrew-driven stage using a single unit of an optical encoder. In: Proc. 26th ASPE annual meeting
8. Lee CB, Kim G-H, Lee S-K (2012) Uncertainty investigation of grating interferometry in six-degree-of-freedom motion error measurements. Int J Precis Eng Manuf 13(9):1509–1515
9. Lee CB, Lee S-K (2012) Measurement of a six-degree-of-freedom dynamic characteristics using angle sensor-implemented grating interferometry. J Korean Soc Precis Eng 29:81–87 (in Korean)
10. Kim W-S, Lee D-J, Lee S-K (2011) Influence of a high vaccum on the precise positioning using an ultrasonic linear motor. Rev Sci Instrum 82:1015112
11. Kang J-H, Lee CB, Joo J-Y, Lee S-K (2010) Phase-locked loop based on machine surface topography measurement using lensed fibers. Appl Opt 50(4):460–467
12. Smith S (2002) Flexures: elements of elastic mechanisms. Taylor & Francis, London
13. Aphale SS, Devasia S, Reza Moheimani SO (2008) High-bandwidth control of a piezoelectric nanopositioning stage in the presence of plant uncertainties. Nanotechnology 19:125503
14. Yong YK, Moheimani SOR, Kenton BJ, Leang KK (2012) Invited review article: high-speed flexure-guided nanopositioning: mechanical design and control issues. Rev Sci Instrum 83:121101

15. Parmar G, Barton K, Awtar S (2014) Large dynamic range nanopositioning using iterative learning control. Precis Eng 38(1):48–56
16. Lee CB, Lee S-K, Tarbutton JA (2014) Novel design and sensitivity analysis of displacement measurement system utilizing knife edge diffraction for nanopositioning stages. Rev Sci Instrum 85:095113
17. Lee CB, Tarbutton JA (2015) Compliance and control characteristics of an additive manufactured-flexure stage. Rev Sci Instrum 86(045107)
18. Lee CB, Lee S-K, Tarbutton JA (2015) Positioning control effectiveness of optical knife edge displacement sensor-embedded monolithic precision stage. Sensors Actuators A 233:290–296
19. Chae KW, Kim W-B, Jeong YH (2011) A transparent polymeric flexure-hinge nanopositioner, actuated by a piezoelectric stack actuator. Nanotechnology 22:335501
20. Hoover AM, Avadhanula S, Groff RE, Fearing RS (2006) A rapidly prototyped 2-axis positioning stage for microassembly using large displacement compliant mechanism. In: Proc. 2006 IEEE international conference on robotics and automation, Orlando, FL, May 2006
21. Hecht E (1987) Optics, 4th edn. Addison-Wesley Longman
22. Born M, Wolf E (1980) Principles of optics: electromagnetic theory of propagation, interference, and diffraction of light, 6th edn. Pergamon Press, New York
23. Collin RE (1985) Antennas and radiowave propagation. McGraw-Hill, New York
24. Ishimaru A (1997) Wavepropagation and scattering in random media. IEEE Press, Piscataway
25. Giovaneli CL (1984) An analysis of simplified solution for multiple knife-edge diffraction. IEEE Trans Antennas Propag 32(3):297–301
26. Davis BA, Brown GS (2002) Diffraction by a randomly rough knife edge. IEEE Trans Antennas Propag 50(12):1769–1778
27. Bennett HE, Porteus JO (1961) Relation between surface roughness and specular reflectance at normal incidence. J Opt Soc Am 51(2):123–129
28. Bhushan B (2013) Introduction to tribology. Wiley
29. Saito Y, Arai Y, Gao W (2009) Detection of three-axis angles by an optical sensor. Sens Actuator A 150:175–183
30. (2014) Micropositioning, nanopositioning, nanoautomation solutions for cutting-edge technologies. Physik Instrumente Catalog. http://www.pi-usa.us/pdf/2011_PICatalog.pdf. Accessed 21 Nov 2014
31. Yong YK, Aphale SS, Reza Moheimani SO (2009) Design, identification, and control of a flexure-based XY stage for fast nanoscale positioning. IEEE Trans Nanotechnol 8(1):46–54
32. Kim Y-S, Dagalakis NG, Gupta SK (2014) Design of MEMS based three-axis motion stage by incorporating a nested structure. J Micromech Microeng 24:075009
33. Sun L, Wang J, Rong W, Li X, Bao H (2008) A silicon integrated micro nano-positioning XY-stage for nano-manipulation. J Micromech Microeng 18:125004
34. Upcraft S, Fletcher R (2003) The rapid prototyping technologies. Assem Autom 23(4):318–330
35. Hague R, Mansour S, Saleh N, Harris R (2004) Materials analysis of stereolithography resins for use in rapid manufacturing. J Mater Sci 39(7):2457–2464
36. Stampfl J, Baudis S, Heller C, Liska R, Neumeister A, Kling R, Ostendorf A, Spitzbart M (2008) Photopolymers with tunable mechanical properties processed by laser-based high resolution stereolithography. J Micromech Microeng 18:125004
37. Yong YK, Lu T-F, Handley DC (2008) Review of circular flexure hinge design equations and derivation of empirical formulations. Precis Eng 32:63–70
38. Tseytlin YM (2002) Notch flexure hinges: an effective theory. Rev Sci Instrum 73(9):3363–3368
39. Lobontiu N, Paine JSN, Garcia E, Goldfarb M (2001) Corner-filleted flexure hinges. J Mech Des 123:346–352
40. Paros JM, Weisbord L (1965) How to design flexure hinges. Mach Des 37:151–156
41. Ling C-B (1968) On stress-concentration factor in a notched strip. J Appl Mech 35(4):833–835
42. Peterson RE (1974) Stress concentration factors. Wiley, New York

43. (2014) Material information: VisiJet® SL tough. 3D Systems. http://www.3dsystems.com/sites/www.3dsystems.com/files/projet_6000_7000_0514_us_web.pdf. Accessed 6 Aug 2014

44. Shilpiekandula V, Youcef-Toumi K (2010) Dynamic modeling and performance trade-offs in flexure-based positioning and alignment systems, motion control. In: Casolo F (ed). InTech. ISBN: 978-953-7619-55-8. doi:10.5772/6981. http://www.intechopen.com/books/motion-control/dynamic-modeling-and-performance-trade-offs-in-flexure-based-positioning-and-alignment-systems

45. Mahmoodi SN, Jalili N (2008) Coupled flexural-torsional nonlinear vibrations of piezoelectrically actuated microcantilevers with application to friction force microscopy. J Vib Acoust 130:061003

46. Braunsmann C, Prucker V, Schäffer TE (2014) Optical-knife-edge displacement sensor for high-speed atomic force microscopy. Appl Phys Lett 104:103101

47. Karabacak D, Kouh T, Huang CC, Ekinci KL (2006) Optical knife-edge technique for nanomechanical displacement detection. Appl Phys Lett 88:193122

48. Lee CB, Lee S-K, Tarbutton JA (2015) Long-term positioning effectiveness of additive manufactured-monolithic double compound notch type flexure stage. In: ASPE 2015 spring topical meeting, Raleigh, NC, 26–29 April 2015

Chapter 23
Microfluidics for Mass Measurement of Miniature Object Like Single Cell and Single MicroParticle

Md. Habibur Rahman, Mohd Ridzuan Ahmad, Masaru Takeuchi, Masahiro Nakajima, Yasuhisa Hasegawa, and Toshio Fukuda

Abstract Single Cell Mass (SCM) is an intrinsic property of a single cell; it arouses a great interest among scientists as cell mass depends on the synthesis of proteins, DNA replication, cell wall stiffness, cell cytoplasm density, cell growth, ribosome, and other analogues of organisms. To date, several great strides have been taken to the advancements of SCM measurement techniques. Nevertheless, more works are required to enable the technology to push frontier in deep analysis of SCM measurement, hence to elucidate intracellular properties. In this book chapter, we present a Lab-on-Chip microfluidics system for SCM measurement, related to the force required to drag a single cell and Newton's law of motion inside microfluidics channel. The drag force on the cell was generated by a pressure-driven syringe micropump, and the motion of the cell was measured using optical observation under an inverted microscope. This approach of measuring SCM was calibrated using known mass (77.3 pg) of a polystyrene microparticle of 5.2 μm diameter. Furthermore, we used *Saccharomyces cerevisiae* baker's yeast cells of different

M.H. Rahman
Department of Computer Science and Engineering, University of Asia Pacific, Green Road, Dhaka-1215, Bangladesh

M.R. Ahmad (✉)
Department of Control and Mechatronics Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
e-mail: ridzuan@fke.utm.my

M. Takeuchi • Y. Hasegawa
Department of Micro-Nano Systems Engineering, Nagoya University, Nagoya, Aichi, Japan

M. Nakajima
Center for Micro-nano Mechatronics, Nagoya University, Nagoya, Aichi, Japan

T. Fukuda
Institute for Advanced Research, Nagoya University, Nagoya, Aichi, Japan

Faculty of Science and Engineering, Meijo University, Nagoya, Aichi, Japan

Intelligent Robotics Institute, School of Mechatronic Engineering, Beijing Institute of Technology, Beijing, China

sizes (2–7 μm diameter) for SCM measurement. Mass of 4.4 μm diameter of single yeast cell was measured as 2.12 pg which is in the range of previously reported single yeast cell mass (2–3 pg). In addition, we also studied the relation between SCM and single cell size. Results showed that single yeast cell mass increases exponentially with the increasing of single cell size.

Keywords Microfluidics • Mass measurement • Single cell analysis • Drag force • Newton's law of motion

1   Introduction

Micro Electro Mechanical Systems (MEMS) provide an excellent platform to analyze single cell mechanics often known as Lab-on-Chip (LoC) microfluidic devices [1–5]. Studies on single cell mechanics acquire a great interest of scientists as cell mechanics can be related to the early diagnosis of diseases from the variations of cells fundamental properties such as cell stiffness and cell wall strength [6]. Cell mechanics consist of (but not limited to) cell wall strength, cell mass, density, and volume at different phases of cell growth cycle. Among them, Single Cell Mass (SCM) is an important parameter as cell mass depends on the synthesis of proteins, DNA replication, cell wall stiffness, cell cytoplasm density, cell growth, ribosome, and other analogous of organisms [7].

Cell mass depends on the synthesis of proteins, DNA replication, cell wall stiffness, cell cytoplasm density, cell growth, ribosome, and other analogous of organisms [7]. Chronic diseases like cancer and tumor affect intracellular physiological properties of cells [8], subsequently cell mass and density will be changed as well [9, 10]. Single cell mass can be measured either from an absolute single cell or from a large population of cells. But from the average data of single cell mass, it is not possible to identify the mass of a particular single cell. Hence, ambiguities arise in terms of cells physical condition's data.

On the other hand, an absolute single cell mass data is able to explain the physical conditions of a particular cell. This leads to differentiate an unhealthy cell from a bunch of healthy cells and vice versa for the healthy cell. Figure 23.1 shows a concept, how single cell mass contributes in terms of identifying infected cell. Figure 23.1a describes a schematic diagram of a population cell. From the estimated data of population cell mass, it is not possible to differentiate the healthy and infected cell. On the other hand, Fig. 23.1b shows the single cell property for a particular cell only. For an infected cell, all these internal particles are being affected by the foreign agents or materials. These physiological changes also affect the mass of single cell. For example, in a tumor infected cell, integrity of DNA faces continuous challenges and genomic instability occurs to the chromosome's structure [11]. Inevitably, this will cause severe change to DNA replication, cytoplasm density, and cell volume which ultimately leads to the changes in single cell mass. In this condition, if we could determine the mass of a single cell, we will be able
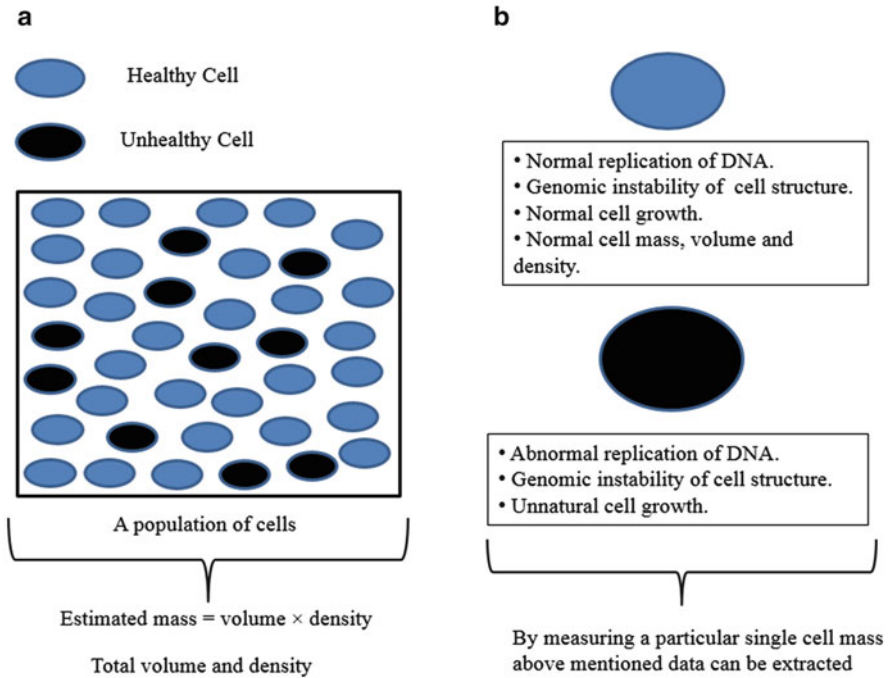
Fig. 23.1 (**a**) A large population of cells. (**b**) Properties of particular healthy and unhealthy cell

to differentiate the unhealthy cell from healthy cells by investigating single cell mass property. Single whole cell mass has also a great contribution in terms by generating biomarkers for rapid identification of intact microorganisms like virus and bacteria [12].

As a result, we strongly believe that studying single cell mass and its measurement techniques will enhance our understanding of physiological properties of cell and perhaps it may provide new tools to diagnosis disease through the variation of single cell mass property of identical cells at different health conditions.

1.1 Technological Advancement of Single Cell Mass Measurement System

With the revolution of micro-bio and nano-bio technology, physiology of single cells is being discovered day by day. Great strides have been taken to develop the technology to investigate the intracellular and extracellular properties of single cell. For example, analysis of single cell inside environmental scanning electron microscope (ESEM) [13–16], AFM cantilever for single cell strength analysis [17], nanoscale electrochemical probe for single cell analysis (SCA) [18], SCA through

electrochemical detection [7, 14, 18–22], and microfluidics disk for single cell viability detection [23]. SCA elucidates complex cellular functions such as cell's mechanical, electrical, and chemical properties. Single cell mechanics is one of the vital parts of the single cell analysis.

Lab-on-chip integrated with microfluidics system enabled scientist to measure the mass of individual cells directly (in the presence of alive cells). Micro-nano mechanical resonators have opened the doors for single cell mass measuring with high accuracy. Frequency of the resonator is inversely proportional to the acquainted mass of resonator [17, 24–28]. Using this principle Burg et al. proposed suspended microchannel resonator (SMR) for single cell mass measurement [4, 19]. But this work was limited to dry cell only, i.e., nonadherent yeast cell [29]; infected microbeads were used to characterize the sensor [30]. This issue has been overcome by "living cantilever arrays"' for measuring adherent HeLa cell [7]. However, resonating cantilever has nonuniform mass sensing ability [31]; as a result, accuracy depends on the cell position merely [7]. On the other hand, cantilever-based mass measurement techniques were not able to relate the cell mass with cell growth through cell cycle. These issues have been addressed by object position-independent pedestal mass measurement [2].

Recently, an excellent method has been proposed for rapid measurement of single cell mass using an Optically Induced Electrokinetics (OEK) microfluidics platform [32]. A time-controlled projecting light pattern was used to illuminate the designated area of the chip to lift cells from surface to a desired height. By competing the buoyancy, gravitational, and drag force acting on a single cell, SCM was estimated. Even though this approach illustrated the possibility of rapid measurement of single cell mass, this work was not able to describe the relationship between single cell mass and cell growth. Consequently, the system will be required to recalibrate each time for measuring different sizes of single cell mass. Therefore, an alternative yet competitive method is necessary to enable the technology to push frontier in SCM measurement, hence single cell mechanics. In response to this challenge, we proposed a Lab-on-Chip microfluidics system for single cell mass measurement rapidly and accurately. However, earlier version of this work has been recently published in an archived journal [33, 34]. This book chapter presented more details in experimental procedures and highlighted the contributions of this work to the community.

2 Proposed Concept: Drag Force for Single Cell Mass Measurement

In this method, single cells were flown through the microfluidics channel and force required (drag force) to move a cell was measured using the parameters that acting along to provide drag force on cell. Then, calculated drag force was related with Newton's law of motion to measure the mass of a single cell. Our proposed method requires a simple microfluidics chip with general laboratory facility like microscope
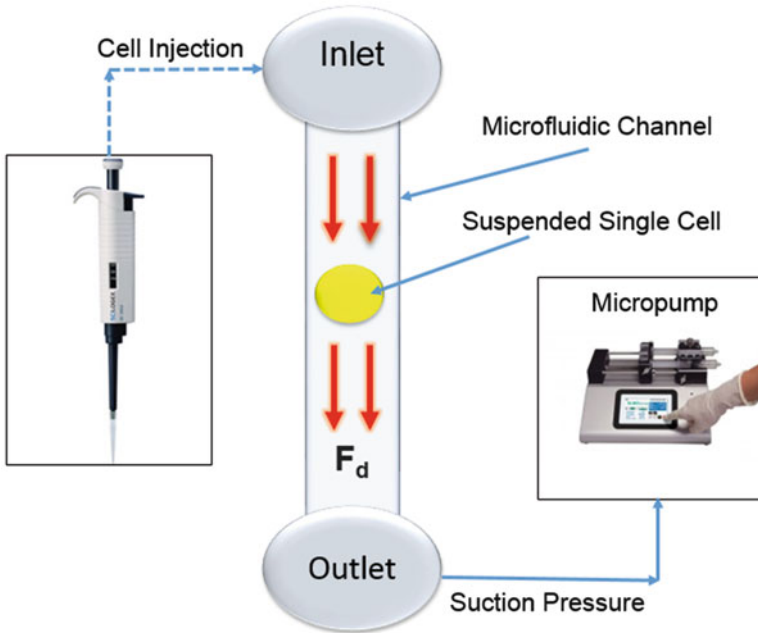
Fig. 23.2 Microfluidics channel for single cell mass measurement. Suction pressure has been applied to outlet of the channel which generates drag to the cell. This drag force has been related with Newton force of motion to measure the mass of single cell

with camera, micropump, and computational system. Figure 23.2 illustrated this concept of single cell mass measurement using proposed microfluidics system. Firstly, the proposed system has been calibrated using known mass of single polystyrene microbeads, and then single yeast cell mass was measured.

A pressure-driven micropump was used to control the flow of the liquid inside the microfluidic channel. During the fluidic flow through the microfluidic channel, along with the liquid flow streamline, it also applies drag force to the cells present inside the channel [35]. Drag force applied on the cell depends on the cell's velocity and surface area for a constant density of medium. This relationship illustrated in Eq. (23.1).

$$F_\mathrm{d} = \frac{1}{2}\rho v^2 C_\mathrm{d} A \qquad (23.1)$$

where F_d is the drag force, ρ is the liquid density, v is the cell velocity, C_d and A is the cell's drag coefficient and area, respectively. Among the above parameters, velocity and cell's area can be measured using optical observation under microscope. Drag coefficient is a dimensionless parameter, depends on the particle's geometrical shape. For a microfluidics channel, Reynolds number is very low ($R_e \ll 1$), i.e., the flow is fully laminar [36]. At this low Reynolds number, drag coefficient has been

suggested for spherical object as in between 0.1 and 0.5 [19, 20]. To drag a particle inside microfluidics channel, two forces are required, firstly the Newtonian force (F_N) and then the gravitational force (F_g) as in Eq. (23.2).

$$F_d = F_N + F_g \qquad (23.2)$$

Due to the exerted drag force on the cell, cell will accelerate and will move forward. In our experimental conditions, i.e., at a very low Reynolds number the effect of gravity to the single cell is very low, and it can be diminished (i.e., $F_g = 0$) [38, 39]. As a result, drag force required to accelerate the particle can be equated with Newtonian force as in Eq. (23.3).

$$F_d = F_N = ma \qquad (23.3)$$

where F_N is the exerted force, m is the mass of cell, and a is acceleration due to the force. During liquid flow, the microfluidic channel is air tight and the pressure inside the microchannel is fully saturated [40]. Consequently applied pressure inside the channel will be distributed uniformly through the entire microfluidic channel. At this stage, drag force exerted on cell surface is equal to force required to accelerate the cell [41]. Hence, equating Eqs. (23.1) and (23.3) mass of the cell/particle can be measured. Other forces like particle to particle aggregation, collision between cells to channel wall can be neglected in our measurement principle [32].

2.1 Design of the Microfluidic Chip

Design of the microfluidics channel was developed using widely used software SolidWorks 2012. Microfluidic chip was designed to maintain the single cell continuous and smooth inside the microfluidic channel. Entire device has two parts; one is the microfluidics chip and another is the glass surface where the channel is attached. Figure 23.3a shows the overview of the LOC microfluidics system. Dimension of the chip is 3 cm × 1.2 cm × 5 mm. This dimensions are standard for the miniature devices, there would be some variations in the overall dimensions. Figure 23.3b shows the top view of the microfluidic chip. Upper surface is a plane surface except inlet and outlet hole. Thickness of the chip is 5 mm (5000 μm) with an inlet and outlet hole of 1.2 mm. Inlet of the diameter has been designed so that 1 mm diameter of silicon pipe can be inserted directly without any additional device. The microfluidics channel is situated at the bottom of the microfluidic chip. Length and width of the microfluidic channel is 20 mm and 15 μm, respectively. The length of the microfluidics channel can be varied according to the purpose of the work. In general, width of the microfluidics channel varied from 10 to 100 μm, depending on the sample cell sizes.

In this work, yeast cells were used as a sample cell and the diameter of the cell varied from 3 to 6 μm. It is suggested that the microfluidics channel width could be in the range of 10–20 μm for single yeast cells flow [42]. To realize this issue,
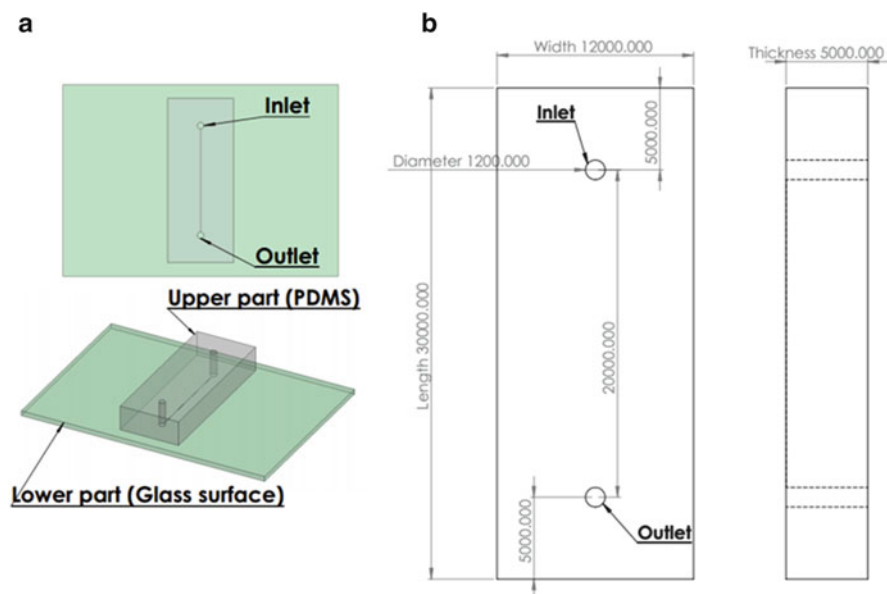
Fig. 23.3 (**a**) LOC microfluidic system. It has two parts: the lower part is the glass surface and the upper is the PDMS microfluidic chip. (**b**) *Top view* of the microfluidic system. All the dimensions are in micrometer. The thickness of the chip 5 mm (5000 μm). The *inlet* and *outlet* hole is 1.2 mm each

we performed several finite element simulation and finally the optimum width was chosen as 15 μm. At this width of the microfluidics channel, single yeast cells maintain their flow properly, nevertheless slight variations does effect the results of single cell mass measurement. Width of the proposed microfluidics channel is 10 μm. When a single cell flows through the microfluidics channel, cells remain suspended at the 70 % of its total volume due to the buoyant force. As a result, in this work height of the channel was designed so that at a time a single cell only flows through the channel. The proposed microfluidics channel is not hanging rather placed on the glass surface; as a result, aspect ratio of the channel to the width didn't effect to the measurement. Figures 23.4 and 23.5 depict the dimensions and 3D part of the microfluidic chip.

2.2   Fabrication of the Microfluidic Chip

LOC microfluidics system was fabricated using polydimethylsiloxane (PDMS, SILPOT 184, Dow Corning Corp.) material. PDMS is a transparent and biocompatible material, which makes this material very popular for biomedical applications. Figure 23.6a, b describes the detail of the procedures that we used for fabrication.

Fig. 23.4 *Bottom view* of the microfluidic chip. The channel is situated at the *bottom* of the part. The total length of the channel is 20 mm, and the width of the channel is 15 µm
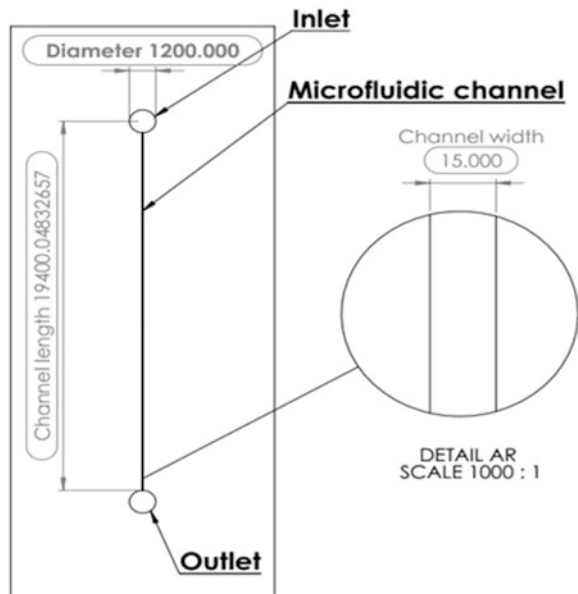
Fig. 23.5 Depth of the microfluidic channel is 10 µm, and the channel depth is uniform
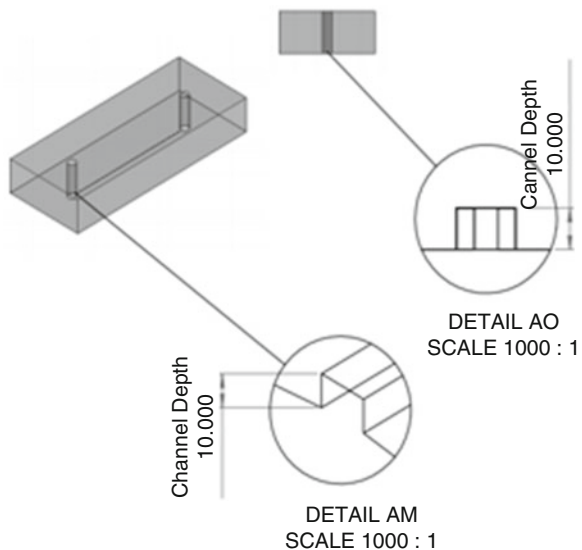
We developed the master mold on silicon surface using soft lithography technique. Figure 23.6a shows the schematic of master mold. Width and depth of the channel is 15 µm and 10 µm, respectively. PDMS prepolymer material was poured on the mold surface and cured for 24 h at the room temperature. Then, the PDMS replica was pilled of from the surface and drilled the inlet and outlet hole. Diameter of the inlet and outlet is 1 mm. Finally, the PDMS chip is ready to use (Fig. 23.6c). In
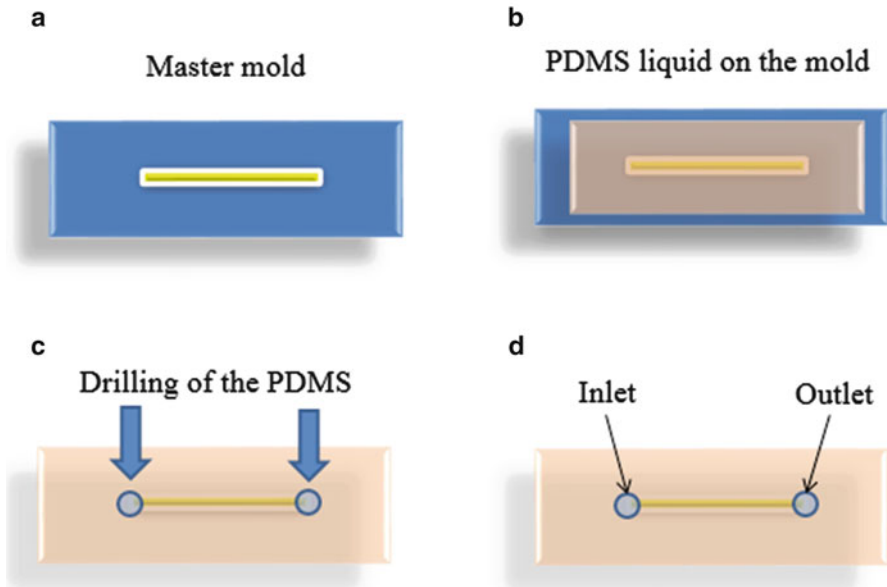
Fig. 23.6 Fabrication procedures of the PDMS microfluidic channel. (**a**) Master mold after soft photolithography. (**b**) PDMS liquid layer on the master mold. (**c**) Dried PDMS structure and drilling of the channel. (**d**) *Inlet* and *outlet* of the microfluidics channel

Fig. 23.6d, the PDMS is placed on the glass surface and ready for the further testing. Figure 23.7a 3D view of the microfluidics channel, (B) shows the top view of the channel and (C) shows the depth of the microfluidics channel.

   Lab-on-chip microfluidics system was fabricated in Micronano System Engineering Laboratory, Nagoya University, Japan. Other than microfluidic chip fabrication, we performed all the experiments in our laboratory. The experiment was conducted under inverted microscope, IX73, *Olympus*. Legato 200, syringe micropump (*KdScientific*) was used to control the particle flow inside microfluidic channel, and powerful computing system was used to perform finite element and data analysis of single cell mass measurement. Figure 23.8 shows a typical protocol of our experimental setup. Inlet shows the fabricated microfluidic chip.

2.3   Water Flow Through Microfluidic Channel

Once the microfluidic channel is fabricated and assembled with the glass surface, we have tested the flow of liquid through the microfluidic channel. Since, inside the microfluidics channel the Reynolds number is very low ($R_e \ll 1$) flow of water inside the channel is fully laminar [43]. We have used water as sample liquid
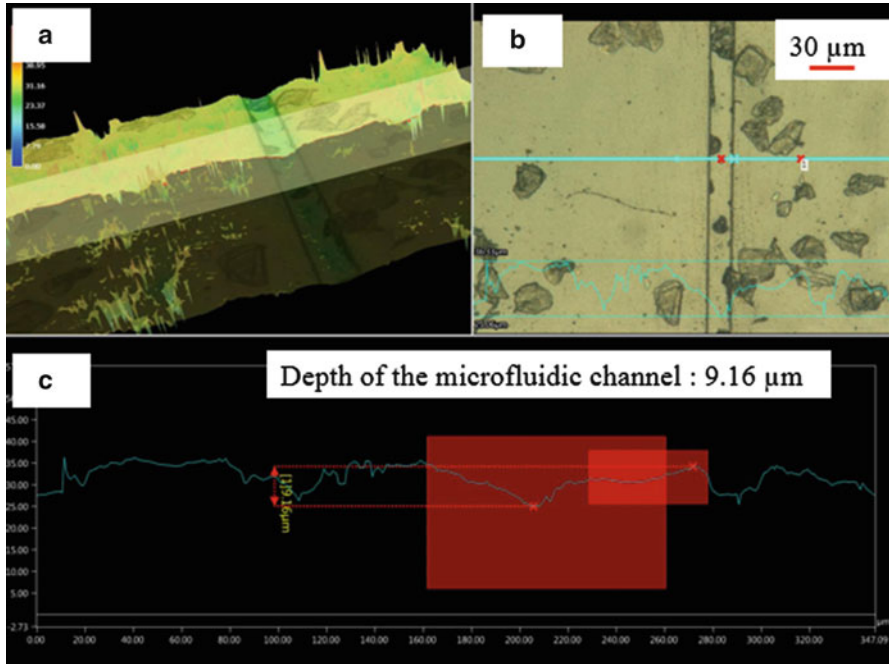
Fig. 23.7 Fabricated PDMS microfluidics system. (**a**) 3D view of the microfluidics channel. (**b**) *Top view* of the channel. (**c**) Depth of the microfluidics channel is 9.6 μm (approximately 10 μm). The images have been captured using Keyance Digital Microscope: VHX 5000
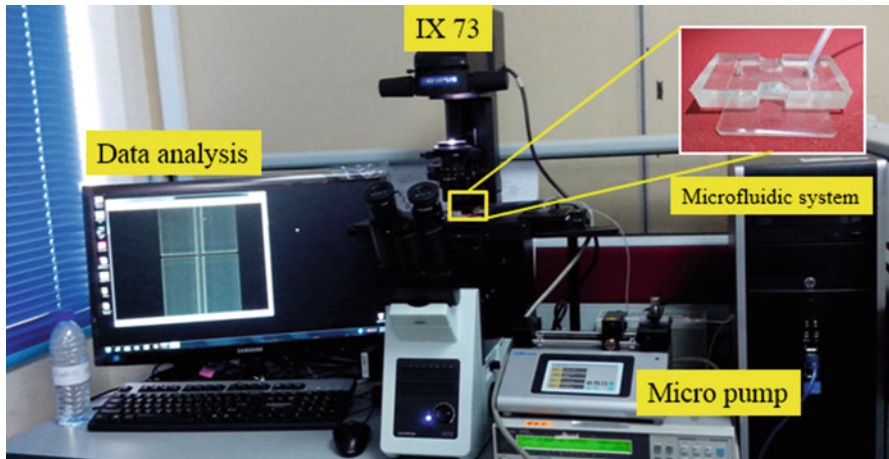
Fig. 23.8 Experimental setup of lab-on-chip microfluidic system for single cell mass measurement
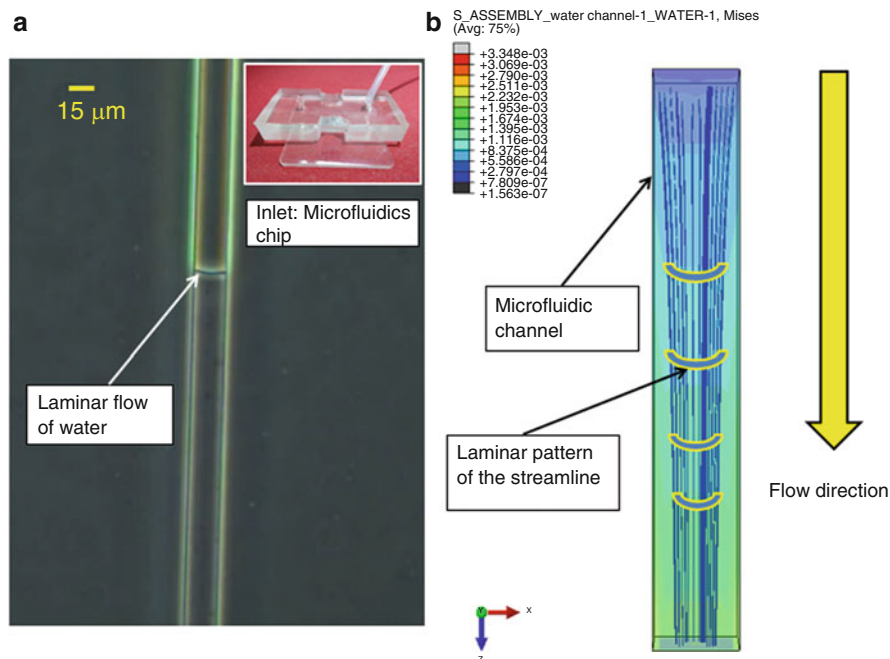
Fig. 23.9 Water flow through the microfluidic channel. (**a**) Water maintains the laminar pattern inside the microfluidic channel. (**b**) FE analysis of the water flow. Streamline of the flow illustrates that maximum pressure of the liquid occurred at the center streamline

and observed the flow pattern inside microfluidic channel. Figure 23.9a shows the water flow through the microfluidic channel. It is depicted that flow of water is in laminar pattern and the maximum pressured occurred at the center streamline of the microfluidics channel. Meanwhile, we also tested this concept using finite element tools to extract the streamline of the liquid flow. Figure 23.9b shows the streamline of the liquid flow through the microfluidic channel.

3 Calibration of the LOC Microfluidics System for SCM Measurement

This section presents the proposed LOC microfluidics system for single cell mass measurement. We have calibrated PDMS microfluidic mass measurement system using commercially available polystyrene microbeads. Spherotech PP-50-10 polystyrene particles were used to verify the mass measurement approach. As presented earlier, Eqs. (23.1) and (23.3) governed the microfluidics-based mass measurement system. These two equations are recalled in this section to enhance the clarity to the reader.

$$F_d = \frac{1}{2}\rho v^2 C_d A \qquad (23.4)$$

$$F_N = ma \qquad (23.5)$$

Equation (23.1) illustrated the drag force required to induce motion of a single particle while Eq. (23.3) illustrated the force requires to accelerate the particle. Again, inside the microfluidics channel, Reynolds number is very low ($R_e \ll 1$), and the liquid flow maintains its laminar properties; the drag force exerted on particle's surface is equal to force required to accelerate the corresponding particle [41] as illustrated in Eq. (23.6).

$$F_N = F_d = ma \qquad (23.6)$$

From this equation we can directly calculate the mass of the single particle if the acceleration (a) and drag force (F_d) are known. Subsequently, to measure single particle mass (m), two major procedures are taken into account. Firstly, the particle's acceleration measurement which can be measured from the velocity and time lapse. Secondly, drag force measurement for that particle. According to Newton's law of motion, these two parameters are directly related to each other for a particular object and are directly proportional at constant mass of the object. However, prior to measure single cell mass, we calibrated this system with a known mass of commercially available single microbead. As obtained from the particle datasheet, average diameter of the microparticle is 5.2 μm, volume 73.6 μm^3, and the density of the polystyrene is 1050 kg/m^3. From these data, theoretical mass of the single microbeads was calculated as 77.3 fg (femto gram). After calibrating our proposed method, we compared our experimented result with the suggested mass from the product datasheet.

3.1 Measuring the Velocity and Acceleration of the Particle

Measuring the velocity and acceleration of the particle/microbeads is the most challenging part of this experiment. The experiment was conducted under inverted microscope, IX73, Olympus. Legato 200, Syringe micropump (KdScientific) was used to control the particle flow inside microfluidic channel. To measure the velocity and acceleration of the particle, four steps were considered. (a) Sample injection, (b) apply suction pressure at the inlet, (c) record the flow of the particle, finally (d) measure the velocity and acceleration of the particle. Firstly, the particle samples were injected into inlet, and then in second step suction pressure was applied at outlet to generate the liquid flow. The suction pressure will direct the microbeads at the gate of microchannel and microbeads will tend to flow. We have applied withdraw flow rate of 50 μL/min for 10 min. This flow rate generated a pressure of approximately higher than 10 kPa at the inlet. Pressure was measured by T-
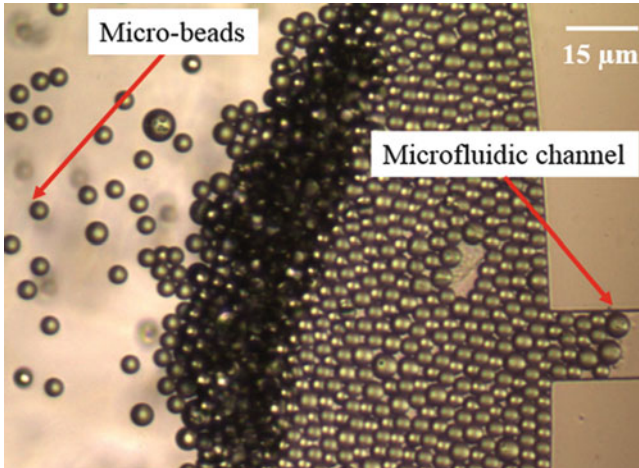
Fig. 23.10 Polystyrene microbeads inside microfluidic channel. Average diameter of the each bead is 5.2 μm

tube connection of MPX10 Piezoresistive pressure sensor, Freescale Semiconductor. According to the Hagen–Poiseuille law, pressure inside microfluidics channel drop significantly [44]. This pressure drop will cause the particles to flow through the microfluidics channel from inlet towards outlet [45]. Figure 23.10 showed that microbeads are tending to flow inside microfluidic channel. Initially, single particle moves very fast due to viscous flow of water.

But at the saturation stage when the channel is full with water, we were able to observe single particle flow clearly (Fig. 23.11a). In third step, when single particle is moving through the microfluidic channel, we have recorded the flow of particle using FastStone video recorder software for 3.2 s (recording time is adjustable). Finally, in fourth step, we have split the 3.2 s video in every 0.2 s and develop a time lapse using our laboratory developed (Matlab Simulink Coding) as shown in Fig. 23.11b. Figure 23.11b illustrates that total distance covered in this time lapse was 120 μm in approximately 3.2 s with the average velocity of 37.5 μm/s. The average acceleration of the particle was measured as 18.56 μm/s^2.

3.2   Characterizing the Mass of Single Microparticle

Once, acceleration of the particle has been measured, we need to measure the force which generated this acceleration. This force is equal to the drag force exerted on the particle. Drag force can be calculated using Eq. (23.1) if the parameters are known. For this experiment, we used normal water as a particle's suspension medium. We have measured the density of the water separately using Shimadzu, *ATX224* balance machine for high accuracy measurement. Density of the water was
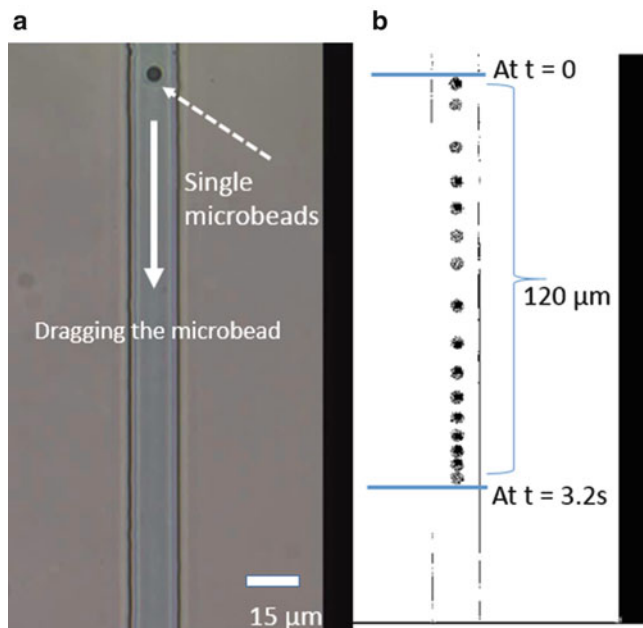
Fig. 23.11 (**a**) Image captured from inverted microscope shows that single microbead is flowing through the microfluidic channel due to drag force exerted on cell from the pressure-driven micropump. (**b**) Time lapse image of the particle flow for 3.2 s. Distance covered in this time is 120 μm. Average velocity of the particle was 37.5 μm/s

measured as 1110 kg/m^3. Area (πr^2) for a 5.2 μm diameter of spherical single microbead is 21.2 μm^2, velocity of the moving particle was measured as 37.5 μm/s, and dimensionless parameter drag coefficient for the spherical particle is 0.1 [37].

Using these parameters, drag force was measured as 1.65×10^{-18} N which generated an acceleration of 18.56 μm/s^2 to the particle. At this stage, pressure-driven drag force is equal to the force of Newton second law of motion which depends on the particle mass [35, 41, 46]. Once the acceleration and the exerted drag force on single cell were measured, equating Eqs. (23.1) and (23.3), the mass of single polystyrene particle was measured 88.9 fg (femto gram). To identify the confidence level, the result was compared with the suggested mass from the particle's datasheet. The suggested mass from particle's datasheet is 77.3 fg which is very consistent to the obtained mass (88.9 fg) using the proposed microfluidics system. From these calibration results, we can envisage that proposed LOC microfluidics system is suitable for single particle/cell mass measurement. To measure single cell mass, instead of using animal cells we have used baker yeast cell as a sample cell for mass measurement.

4 Single Cell Mass Measurement

4.1 Culturing of Yeast Cells

Yeast cell has been cultured using conventional cell culturing methods. Yeast growth medium/agar has been developed using YPD and sucrose in a ratio of 1:1 in one liter of water. The mixture was then shaken for few minutes to ensure the proper dilution of the YPD and sucrose. Later on, commercially available baker's yeast powder was inserted inside the growth medium. The mixture of the YPD, sucrose, and the yeast powder was then kept inside incubator (Constance, *Germany*) at 30 °C temperature. We have cultured cells in three different durations: 6, 12, and 24 h. Different duration of culturing times produce different sizes of yeast cell. Lowest size of cultured yeast cell was 2 μm while the maximum size was 7 μm of diameter. The reason for growing different sizes of cell is to investigate the mass of different sizes of single cell. We have also observed budding yeast cells of different sizes. From the cultured yeast cell, we have pipetted some of the cells into petri dish and observed under microscope.

4.2 Mass Measurement of Single Yeast Cell

To measure the mass of single yeast cell, we have used similar approach as we explained in the calibration section. Firstly, we measured the velocity and acceleration of the moving cell; then, we measured the force required to accelerate the cell. Initially, cultured yeast cells were diluted with water and then injected into the microfluidics channel. Using micropump, we applied suction pressure inside the microfluidic channel. Suction pressure generated drag force on single cells and cells started to flow. For this particular measurement, we have selected a well-visible cell which diameter was 4.4 μm, and the volume of the cell was 44.5 μm^3. Due to the applied suction pressure, cells were traversing a displacement of 107 μm in 5.5 s (Fig. 23.12). Displacement was measured using an image analyzing software (Image J), developed by the National Institute of Health, Japan. The initial velocity of the yeast cell was 19.45 μm/s. Drag coefficient of the yeast cell is 0.1 [47], and the density of the yeast cell cultured medium was measured as 1180 kg/m^3 using weight balance equipment (Shimadzu, *ATX224*).

From these parameters using Eq. (23.1) we measured the drag force to move the cell as 3.39×10^{-19} N. Once the cells start to flow we stopped applying pressure and deceleration occurred to the cell. After 5.5 s, the velocity of the yeast cell was 18.5 μm/s. As a result, deceleration occurred with the magnitude of 0.17 μm/s^2. Now by equating Eqs. (23.1) and (23.3), we measured the mass of single yeast cell as 1.9 pg. This result of mass can be varied with cell sizes and the parameters measurement.

Fig. 23.12 Single yeast cell flow through microfluidic channel. Cell covers a distance of 107 μm in 5.5 s at the average velocity of 19.45 μm/s
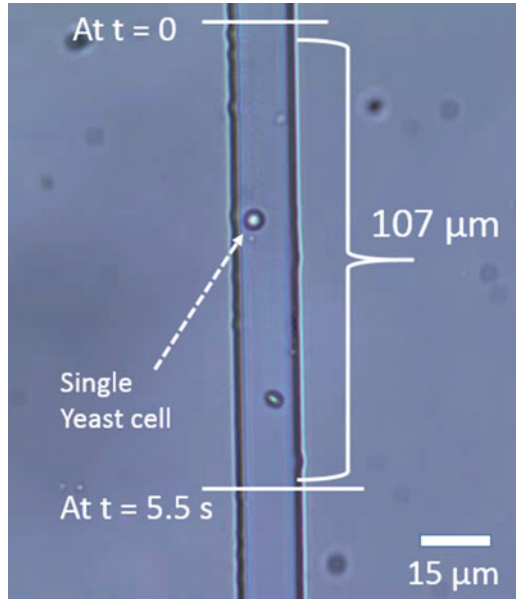
We have tested the repeatability of the system for single cell mass measurement. For a same sample of single cell, we have measured the mass for ten times. Each time the cell position was in different place inside microfluidic channel. Each time flow was controlled using pressure-driven flow. Single cell mass measurement was slightly varied each time, but yet all are very consistent to the reported mass of single yeast cell. Results showed that the measured range of single yeast cell mass was in between 1.4 and 2.7 pg, with the average mass of 2.12 pg. This result is consistent with the cell mass measurement using suspended microchannel resonator (SMR). Previously, reported mass of an adult (4 μm diameter) yeast cell varied from 2 to 3 pg [29]. After verifying the result with the previous reported yeast cell mass, we can say proposed single cell mass measurement system is valid for single cell mass measurement. We have plotted a scattered graph to realize the closeness of the measurement (Fig. 23.13).

4.3 Single Cell Mass vs. Cell Sizes

We have also investigated the relation between cell mass and the radius of the single cells. From our cultured yeast cells we have observed three different sizes of cells. We have measured the mass of the 2.5, 3.5, and 5.5 μm diameter of cells. Figure 23.14 shows the different sizes of yeast cell flowing inside microfluidic channel. Yellow circle is indicating the cell position. Figure 23.14a shows the cell of 2.5 μm diameter, (B) shows the cell of 3.5 μm diameter while (C) shows the cell
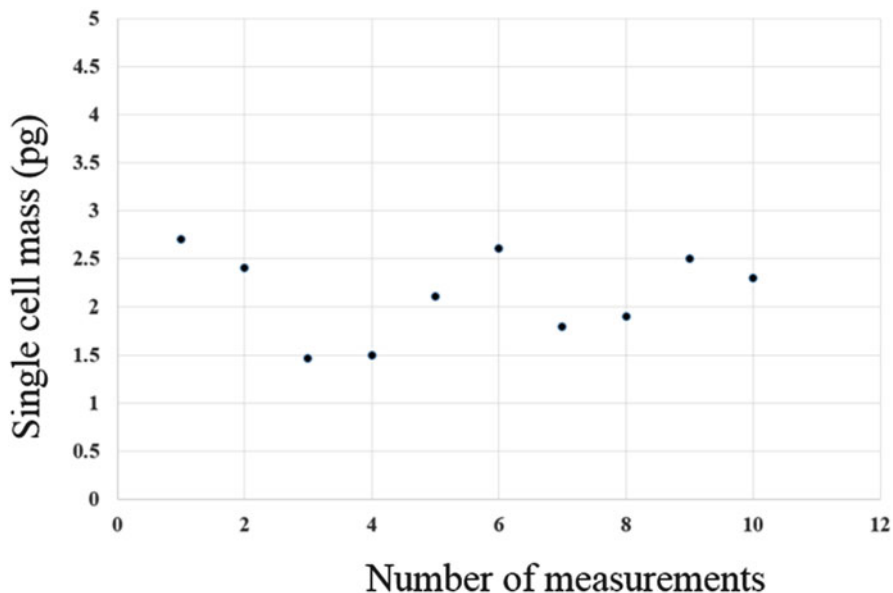
Fig. 23.13 Repeatability of single cell mass measurement. The measurement was conducted ten times in different position inside the microfluidic channel of single yeast cell 4.4 μm diameter. Yeast cell mass measurement varied from 1.4 to 2.7 pg
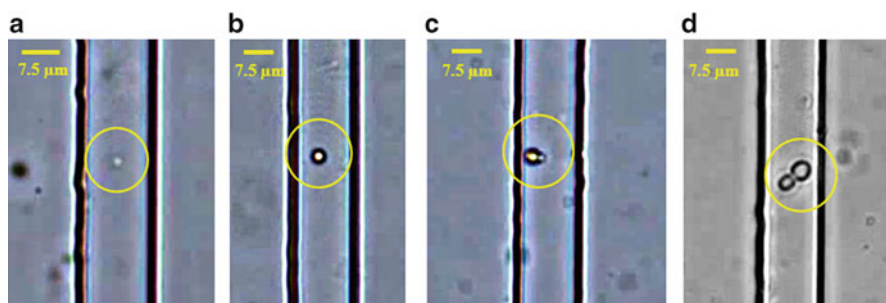
Fig. 23.14 Cultured yeast cell inside microfluidic channel. (**a**) 2.5 μm diameter of yeast. (**b**) 3.5 μm diameter of yeast cell. (**c**) 5.5 μm diameter of yeast cell. *Yellow circle* indicates the cells. (**d**) 7.5 μm diameter of budding yeast cell

with 5.5 μm diameter mother cell budded with 3.5 μm of cell. With the increasing of cell diameter cell volume increases, subsequently causing the mass to be increased as well. We measured the mass of each cells individually. For a diameter of 2.5 μm, 3.5 μm, 5.5 μm, and 7.5 μm, mass of the single cells were measured as 0.9 pg, 1.4 pg, 3 pg, and 6.4 pg, respectively. Figure 23.15 shows the trend of single yeast cell mass increases exponentially with the increasing of the yeast cell sizes. The reported relation between single cell mass and cell size is in agreement with the previous argument [2, 48].
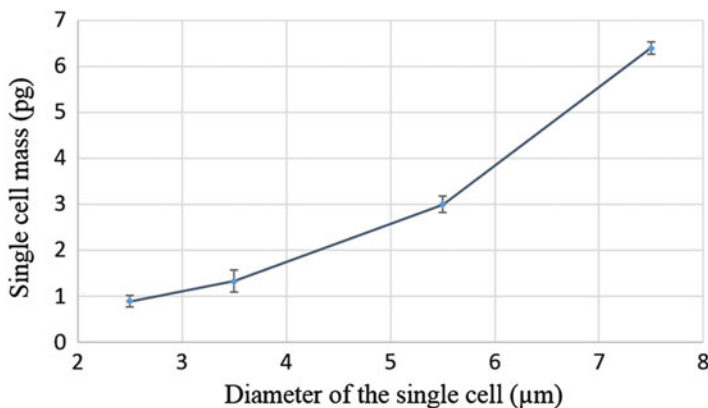
Fig. 23.15 Single cell mass vs. diameter of the single cell. We have experienced an exponential increase in cell mass with increases of cell diameter. *Error bar* shows the standard deviation of the results for ten number measurements. The last most point shows the mass of the budding yeast cell, where diameter of the two budded cell were 7.5 μm

4.4 Validation of the Measurement

Single polystyrene particle's mass measurement results were verified with the suggested mass from the characteristic of the particle (Spherotech PP-50-10) and obtained results for 5.2 μm diameter of a single particle is 87 % closer to the suggested mass of the particle. This variation occurs because datasheet reports only the average mass of the single particle while in this work we measured the absolute mass of single particle. On the other hand, single yeast cell mass results have been verified with previously weighted single yeast cell mass using SMR where mass of single yeast cell (5 μm diameter) varied in between 2 and 4 pg [14] while in our measurement it varies between 2.8 and 3.4 pg for the diameter of 5.5 μm which is up to 95 % close to the reported results. This slight difference could be occurred due to sample preparations, sample sizes, and the suspension medium. However, in our proposed system, there is no use of the frequency shifting phenomenon; as a result, there are no issues of nonuniform mass sensitivity. In addition, drag force-based microfluidics system eases the tedious measurement steps required by SMR but yet provides competitive results of single cell mass. Hence, this method can be useful for rapid and high-throughput measurement of single cell mass. Table 23.1 showed summary of the results and closeness with previously reported single particle/cell mass. However, the accuracy of the results can be improved by measuring required parameters like density of the liquid, velocity and acceleration profiling, and particle size measurement more carefully.

Table 23.1 Summary of the results and closeness with previously reported single particle/cell mass

Sample	Measured mass (g)	Previously reported mass (g)	Closeness
Polystyrene microbead (*Spherotech microparticles*)	88.9×10^{-15}	77.3×10^{-15}	87 %
Single yeast cell (4.5 μm diameter)	1.4×10^{-12} to 2.7×10^{-12}	2×10^{-12} to 3×10^{-12} [29]	70–95 %
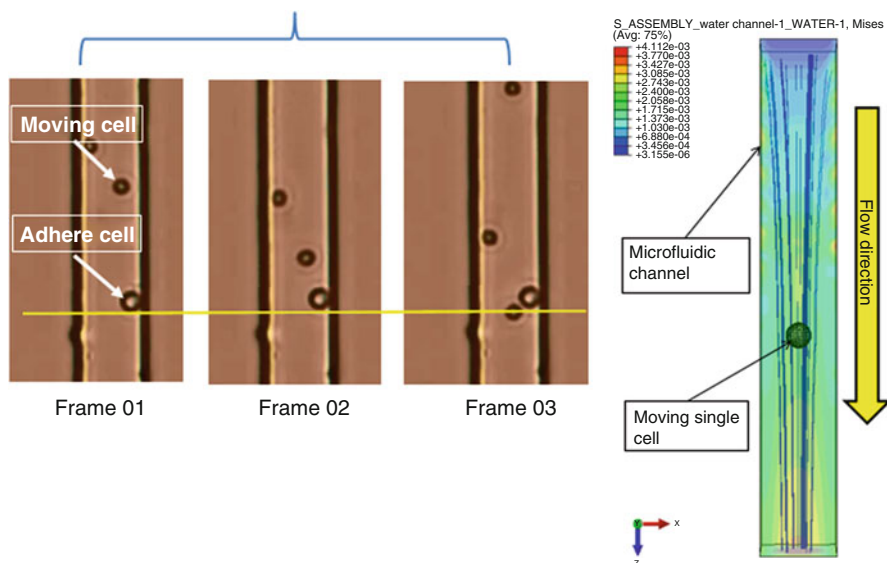
Fig. 23.16 (**a**) Frame 01–03 where adhered cell and moving cell. Moving cell crosses the adherent cell in time. (**b**) Single cell flows through the center streamline of the microfluidic channel

4.5 Effects of the Streamline of the Microfluidic Flow for SCM Measurement

In this work, the flow of the microfluidic channel has a major effect in terms of velocity and acceleration profiling of the particle. Earlier, in Fig. 23.9 it was illustrated that liquid follows the laminar flow profile inside a microfluidic channel, and velocity is maximum at the center streamline of the channel. As a result, it is important to make sure that single cells are in the center of the microfluidic channel and the flow is smooth. On the other hand, if cells become adhered to the channel surface/wall, then it is not be possible to generate the flow to the cell. Figure 23.16a depicts the status of cell adherence in the flow. Frame 01–03 illustrated that moving cells are passing over the adherent cell. While Fig. 23.16b showed the FE analysis of the single cell flow at the center of the streamline. So,

velocity and acceleration profiling must be completed before cells become adhered (approximately in 45 min after sample injection). However, adherent cell can be washed using chemical reagent. In this experiments, adhered cells have been washed using trypsin [7].

4.6 Effect of the Cells Geometrical Shape to the Measurement

Geometrical shape of the sample cell effects the drag coefficient measurement of the particle. In this work, we used yeast cell as sample, and the drag coefficient of the yeast cell has already been reported as 0.8–0.15 [47]. On the other hand, drag coefficient profiling of nonspherical like ellipsoidal or cylindrical particle inside suspended microchannel has been explained by Hensley and Papavassiliou [37]. They have taken great strides to calculate the drag coefficient of ellipsoidal and cylindrical particle. Equation (23.7) shows the drag coefficient for ellipsoidal while Eq. (23.8) shows the drag coefficient of cylindrical particle.

$$\frac{C_{\mathrm{D}}}{C_{\mathrm{D,H\&B}}} = \left[1 + 2.33(\varphi)^{\frac{2}{5}}\left(\frac{d_{\mathrm{p}}}{H}\right)\right] \tag{23.7}$$

$$\frac{C_{\mathrm{D}}}{C_{\mathrm{D,H\&B}}} = \left[1 + 4.20(\varphi)^{\frac{2}{5}}\left(\frac{d_{\mathrm{p}}}{H}\right)\right] \tag{23.8}$$

where φ, d_{p}, H, C_{D}, and $C_{\mathrm{D,H}}$&$_{\mathrm{B}}$ denotes the aspect ratio of particle, diameter of the particle, height of the microduct, drag coefficient of the particle, and drag coefficient measured by Happel and Branner, respectively [37, 49]. If any applications required to measure the mass of elliptical or cylindrical particle, we may use above equations to calculate the drag coefficient. Once drag coefficient is calculated, then we may measure the mass of nonspherical particle using proposed LOC microfluidics system. However, prior to the mass measurement of the particle the drag coefficient is needed to be measured properly.

5 Conclusions

Existing Single Cell Mass (SCM) measurement techniques are limited to frequency shifting phenomenon only and an alternative approach has emerged mandatory to enable the scientists to push frontier in SCM measurement techniques. In this work, a novel SCM measurement method has been proposed, where SCM was measured using drag force inside fabricated lab-on-chip microfluidics system.

A pressure-driven micropump was used to provide drag force inside the microfluidics channel. Exerted drag force accelerated the suspended cells inside the channel and Newton's second law of motion was applied to measure single cell mass.

The proposed approach has been validated by measuring the known mass of commercially available single polystyrene microparticle of 5.2 µm diameter. We also took the attempt to measure mass of single yeast cell; results demonstrated that an adult yeast cell of diameter 4.4 µm has a mass of 2.12 pg. In addition, different sizes (2.5, 3.5, 5.5 µm) of yeast cells mass were measured to study the relationship between cell growth and mass. We observed that single cell mass increased exponentially as cell size increases. The proposed SCM measurement technique eases the existing complex resonant-based single cell mass measurement system. Moreover, this approach brought an alternative yet competitive tool for the single cell mass analyzers; hence, it is envisaged that this work will contribute significantly to the knowledge of single cell mechanics. In the future, this approach can be applied to mass of human cells, and perhaps it may provide new tools for disease diagnosis through the variation of single cell mass property of identical cells in different health conditions.

References

1. Lee J, Chunara R, Shen W et al (2011) Suspended microchannel resonators with piezoresistive sensors. Lab Chip 11:645–651. doi:10.1039/c0lc00447b
2. Park K, Millet LJ, Kim N et al (2010) Measurement of adherent cell mass and growth. Proc Natl Acad Sci U S A 107:20691–20696. doi:10.1073/pnas.1011365107
3. Fritzsch FSO, Dusny C, Frick O, Schmid A (2012) Single-cell analysis in biotechnology, systems biology, and biocatalysis. Annu Rev Chem Biomol Eng 3:129–155. doi:10.1146/annurev-chembioeng-062011-081056
4. Burg TP, Manalis SR (2003) Suspended microchannel resonators for biomolecular detection. Appl Phys Lett 83:2698–2700. doi:10.1063/1.1611625
5. Nilsson J, Evander M, Hammarström B, Laurell T (2009) Review of cell and particle trapping in microfluidic systems. Anal Chim Acta 649:141–157. doi:10.1016/j.aca.2009.07.017
6. Rahman MH, Sulaiman AH, Ahmad MR, Fukuda T (2013) Finite element analysis of single cell wall cutting by piezoelectric-actuated vibrating rigid nanoneedle. IEEE Trans Nanotechnol 12:1158–1168
7. Park K, Jang J, Irimia D et al (2008) "Living cantilever arrays" for characterization of mass of single live cells in fluids. Lab Chip 8:1034–1041
8. De Flora S, Izzotti A, Randerath K et al (1996) DNA adducts and chronic degenerative diseases. Pathogenetic relevance and implications in preventive medicine. Mutat Res Genet Toxicol 366:197–238. doi:10.1016/S0165-1110(96)00043-7
9. Mitchison JM (2005) Single cell studies of the cell cycle and some models. Theor Biol Med Model 2:4–9. doi:10.1186/1742-4682-2-4
10. Cooper S (2006) Distinguishing between linear and exponential cell growth during the division cycle: single-cell studies, cell-culture studies, and the object of cell-cycle research. Theor Biol Med Model 3:10–25. doi:10.1186/1742-4682-3-10
11. Lord CJ, Ashworth A (2012) The DNA damage response and cancer therapy. Nature 481:287–294. doi:10.1038/nature10760

12. Fenselau C, Demirev PA (2002) Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom Rev 20:157–171. doi:10.1002/mas.10004

13. Ahmad MR, Nakajima M, Kojima S et al (2010) Nanoindentation methods to measure viscoelastic properties of single cells using sharp, flat, and buckling tips inside ESEM. IEEE Trans Nanobioscience 9:12–23. doi:10.1109/TNB.2009.2034849

14. Ahmad MR, Nakajima M, Kojima S et al (2008) The effects of cell sizes, environmental conditions, and growth phases on the strength of individual W303 yeast cells inside ESEM. IEEE Trans Nanobioscience 7:185–193. doi:10.1109/TNB.2008.2002281

15. Ahmad MR, Nakajima M, Kojima S et al (2011) Buckling nanoneedle for characterizing single cells mechanics inside environmental SEM. IEEE Trans Nanotechnol 10:226–236

16. Ahmad MR, Nakajima M, Kojima S et al (2008) In situ single cell mechanics characterization of yeast cells using nanoneedles inside environmental SEM. IEEE Trans Nanotechnol 7:607–616

17. Obataya I, Nakamura C, Han S et al (2005) Nanoscale operation of a living cell using an atomic force microscope with a nanoneedle. Nano Lett 5:27–30. doi:10.1021/nl0485399

18. Fasching RJ, Bai SJ, Fabian T, Prinz FB (2006) Nanoscale electrochemical probes for single cell analysis. Microelectron Eng 83:1638–1641. doi:10.1016/j.mee.2006.01.262

19. Burg TP, Godin M, Knudsen SM et al (2007) Weighing of biomolecules, single cells and single nanoparticles in fluid. Nature 446:1066–1069. doi:10.1038/nature05741

20. Cross SE, Jin Y-S, Rao J, Gimzewski JK (2007) Nanomechanical analysis of cells from cancer patients. Nat Nanotechnol 2:780–783

21. Godin M, Bryan AK, Burg TP et al (2007) Measuring the mass, density, and size of particles and cells using a suspended microchannel resonator. Appl Phys Lett 91:123121–1231212. doi:10.1063/1.2789694

22. Xia F, Jin W, Yin X, Fang Z (2005) Single-cell analysis by electrochemical detection with a microfluidic device. J Chromatogr A 1063:227–233. doi:10.1016/j.chroma.2004.12.022

23. Kubo I, Furutani S, Matoba K (2011) Use of a novel microfluidic disk in the analysis of single-cell viability and the application to Jurkat cells. J Biosci Bioeng 112:98–101. doi:10.1016/j.jbiosc.2011.03.016

24. Ilic B, Graighead HG, Krylov S et al (2004) Attogram detection using nanoelectromechanical oscillators. J Appl Phys 95:3694–3703. doi:10.1063/1.1650542

25. Yang YT, Callegari C, Feng XL et al (2006) Zeptogram-scale nanomechanical mass sensing. Nano Lett 6:583–586. doi:10.1021/nl052134m

26. Gupta A, Akin D, Bashir R (2004) Single virus particle mass detection using microresonators with nanoscale thickness. Appl Phys Lett 84:1976–1978. doi:10.1063/1.1667011

27. Lange D, Hagleitner C, Hierlemann A et al (2002) Cantilever arrays on a single chip: mass-sensitive detection of volatile organic compounds. Anal Chem 74:3084–3095

28. Ilic B, Yang Y, Craighead HG (2004) Virus detection using nanoelectromechanical devices. Appl Phys Lett 85:2604–2606

29. Bryan AK, Goranov A, Amon A, Manalis SR (2010) Measurement of mass, density, and volume during the cell cycle of yeast. Proc Natl Acad Sci U S A 107:999–1004. doi:10.1073/pnas.0901851107

30. Weng Y, Delgado FF, Son S et al (2011) Mass sensors with mechanical traps for weighing single cells in different fluids. Lab Chip 11:4174–4180. doi:10.1039/c1lc20736a

31. Dohn S, Sandberg R, Svendsen W, Boisen A (2005) Enhanced functionality of cantilever based mass sensors using higher modes. Appl Phys Lett 86:233501–233503. doi:10.1063/1.1948521

32. Zhao Y, Lai HSS, Zhang G et al (2014) Rapid determination of cell mass and density using digitally controlled electric field in a microfluidic chip. Lab Chip 14:4426–4434. doi:10.1039/C4LC00795F

33. Rahman MH, Ahmad IL, SayahKarajy M, et al (2014) Lab-on-chip microfluidic system for single cell mass measurement. Biomed Eng Sci (IECBES), 2014 IEEE conference 672–676. doi: 10.1109/IECBES.2014.7047590

34. Rahman MH, Ahmad MR, Takeuchi M et al (2015) Single cell mass measurement using drag force inside lab-on-chip microfluidics system. IEEE Trans Nanobioscience 14:927–934

35. Nott PR, Brady JF (2006) Pressure-driven flow of suspensions: simulation and theory. J Fluid Mech 275:157. doi:10.1017/S0022112094002326
36. Martel JM, Toner M (2013) Particle focusing in curved microfluidic channels. Sci Rep 3:1–8. doi:10.1038/srep03340
37. Hensley Z, Papavassiliou D (2014) Drag coefficient correction for spherical and non-spherical particles suspended in square microducts. Ind Eng Chem Res 53:10465–10474
38. Jones AM, Knudsen JG (1961) Drag coefficients at low Reynolds numbers for flow past immersed bodies. AIChE J 7:20–25. doi:10.1002/aic.690070107
39. Moynihan MJ (2013) Double emulsion generation in the mass production of inertial confinement fusion targets using T-junctions. Ph.D. thesis, University of Rochester
40. Fuerstman MJ, Lai A, Thurlow ME et al (2007) The pressure drop along rectangular microchannels containing bubbles. Lab Chip 7:1479–1489. doi:10.1039/b706549c
41. Kim SM, Lee SH, Suh KY (2008) Cell research with physically modified microfluidic channels: a review. Lab Chip 8:1015–1023. doi:10.1039/b800835c
42. Lee J, Bryan AK, Manalis SR (2011) High precision particle mass sensing using microchannel resonators in the second vibration mode. Rev Sci Instrum 82:02370401–02370404. doi:10.1063/1.3534825
43. Kamholz AE, Yager P (2001) Theoretical analysis of molecular diffusion in pressure-driven laminar flow in microfluidic channels. Biophys J 80:155–160. doi:10.1016/S0006-3495(01)76003-1
44. Arlett JL, Roukes ML (2010) Ultimate and practical limits of fluid-based mass detection with suspended microchannel resonators. J Appl Phys 108:084701. doi:10.1063/1.3475151
45. Lee J, Shen W, Payer K et al (2010) Toward attogram mass measurements in solution with suspended nanochannel resonators. Nano Lett 10:2537–2542. doi:10.1021/nl101107u
46. Kubik A, Kleiser L (2004) Forces acting on particles in separated wall-bounded shear flow. Proc Appl Math Mech 513:512–513. doi:10.1002/pamm.200410
47. Wu Y, Sun D, Huang W (2011) Mechanical force characterization in manipulating live cells with optical tweezers. J Biomech 44:741–746. doi:10.1016/j.jbiomech.2010.10.034
48. Mitchison JM (2003) Growth during the cell cycle. Int Rev Cytol 226:165–258. doi:10.1016/S0074-7696(03)01004-0
49. Happel J, Brenner H (1965) Low Reynolds number hydrodynamics. Prentice-Hall, Englewood Cliffs

Chapter 24
Micromanipulation Tools

Jin Li and Zhuming Bi

Abstract This chapter classifies the working principles of micromanipulation and compares the end-effectors developed based on different principles. Differing from the dominating gravitational forces at macroscale, adhesion forces by scaling effect dominate the operations at microscale. One significant challenge is that adhesion forces take time to release objects accurately; new manipulation strategies are required to overcome such a challenge. This chapter also discusses some major manipulation strategies including magnetic levitation, electrostatic levitation, air levitation, and acoustic levitation; these strategies are evaluated using the examples from the literatures.

Micromanipulation refers to the physical interaction of a micromanipulator with a sample. Generally, the level of the required precision of movement in micromanipulation cannot be achieved manually. The dimensional ranges of micromanipulation are a few of micrometers up to millimeters. Micromanipulation is one of the critical techniques in making and assembling microsystems. In particular, it is essential for various micro-assemblies. For example, micromanipulation is widely adopted to assemble parts with different materials; these parts are with complex shapes and from incompatible fabrication processes. Micromanipulation can also be applied in operating miniaturized objects at other phases of their life cycles, and the types

J. Li

Complex and Intelligent System Research Laboratory (CISRL), School of Mechanical and Power Engineering, East China University of Science and Technology, Shanghai 200237, China
e-mail: lijinme@ecust.edu.cn

Z. Bi (✉)

Department of Civil and Mechanical Engineering, Indiana University Purdue University Fort Wayne, Fort Wayne, IN 46805, USA
e-mail: biz@ipfw.edu

of operations can be testing, repairing, or disassembly. In contrast to conventional manipulations, new research issues in micromanipulation are quite different since (1) the manipulated objects are miniaturized and (2) the dominance of different physical phenomena varies during the operation; this affects the interaction between objects. Therefore, both end-effectors and manipulation strategies are fundamentally different between micromanipulation and conventional manipulations.

1 Micromanipulation and Macromanipulation

In this chapter, a *micro-object* refers to an object with its geometric dimensions less than hundreds of micrometers; accordingly, an object whose dimensions are large than hundreds of micrometers is viewed as *macro-object*. Manipulating a micro-object illustrates some distinguished features in comparison with the conventional manipulation on macro-objects [1]:

1. The scaling effect and contact hysteresis cause adhesive forces which dominates the behaviors in micromanipulation
2. The micro-object is very fragile which can be damaged by clamping forces
3. Visual feedback is essential to micromanipulation
4. The end-effector is required to possess high accuracy for micro-motions

The scaling effect [2–4] corresponds to the order change of significance of interaction phenomena when the system changes in its overall size. Due to the scaling effect, adhesion forces dominate the interactions of objects at the microscale, and gravitational forces dominate the interactions of objects whereas at macroscale.

In the micromanipulation, adhesion forces are difficult to be quantified; moreover, such forces are time dependent which take some time to release objects accurately. Manipulated objects tend to adhere to tool or substrate. In other words, adhesion forces bring some uncertainties in manipulating objects at microscale. Therefore, new operating strategies are required to manipulate micro-objects. Note that for macro-objects, gravity forces play the dominating role in the interactions of objects. As a result, releasing and grasping an object typically requires the same movement but with the opposite order. In such a case, the reversibility of handling actions is predictable.

Contact hysteresis [5, 6] means that the force needed to separate two objects in contact with each other is generally larger than the force that is needed to bring them into contact. Contact hysteresis is caused by the change in the contact area between two surfaces; such a change causes contact forces due to the elastic and/or plastic deformations occurring to interacting objects. At the microscale, these deformations are induced by external loads applied on objects. In micromanipulation, external loads are applied by end-effector. Adhesion forces also include the induced forces by creep deformations. In the context of manipulation, contact hysteresis becomes more and more important when the object to be manipulated becomes smaller and smaller.

In comparison with macro-objects, micro-objects could be easily damaged during manipulation; objects can be either deformed or broken. A tool for micromanipulation is generally larger than manipulated objects; it can produce relatively large force than the enduring capacity of the object. Thus, micro-objects are very vulnerable to be damaged, in particular, for the objects with complex shapes and fine details. On the other hand, an end-effector tool is easily damaged as well for the same reasons. Therefore, the manipulation process must be carefully controlled to avoid potential damages on objects or tools. To implement the control, it is critical to obtain the information of both the position and forces related to objects from measurement.

To deal with the unpredictability in handling actions and the difficulty of the measurement on micro-objects, visual feedbacks can be an important source of information to determine the positions and orientations of end-effector tools or manipulated objects. Therefore, microscopes are usually required to be as a part of a manipulation system for micro-objects. However, one issue related to microscopes is that they can confine the workspace of the manipulation tool; using a microscope could affect both of manipulation strategies and the selection of manipulation systems. In addition, data processing with the microscope is another issue; which can demand intensive amount of computation and slow down the manipulation.

When the size of a manipulated object is reduced to a micrometer level, the positional accuracy of the object becomes high. The resolution of measurement is typically required at the sub-micrometer level. Note that the acceptable level of positional accuracy for macro-objects is in a range of a few hundreds of micrometers. For an example of the case when a conventional industrial robot runs in the open-loop control mode. It is clear that the calibration approach of conventional systems for macro-objects [7] is applicable to microsystems due to high-accuracy requirements of the later. New actuation mechanisms and actuator materials are needed for the manipulation on micro-objects.

2   Micro-Handling Principles

The products from micro-objects with a micro-level size are called *micro-products*. They are widely used in two technological fields, i.e., the precision mechanical engineering field and the semi-conduction field. Typical micro-objects are consumer products, opto-electronic components, biomedical precision instrumentation, and microsystems for automotive applications.

A micro-handling principle refers to the physical principle that produces the required force to capture and hold a micro-object at the given position under the reference coordinate system. To fulfill the task, one force that has to be compensated for is the gravitational force of the object. Two other forces occurring to an object are (1) adhesive forces between the environment and the object and (2) inertial forces and moments due to the accelerations when the object moves with the gripper in space. Depending on the object size, the percentage of specific force in the
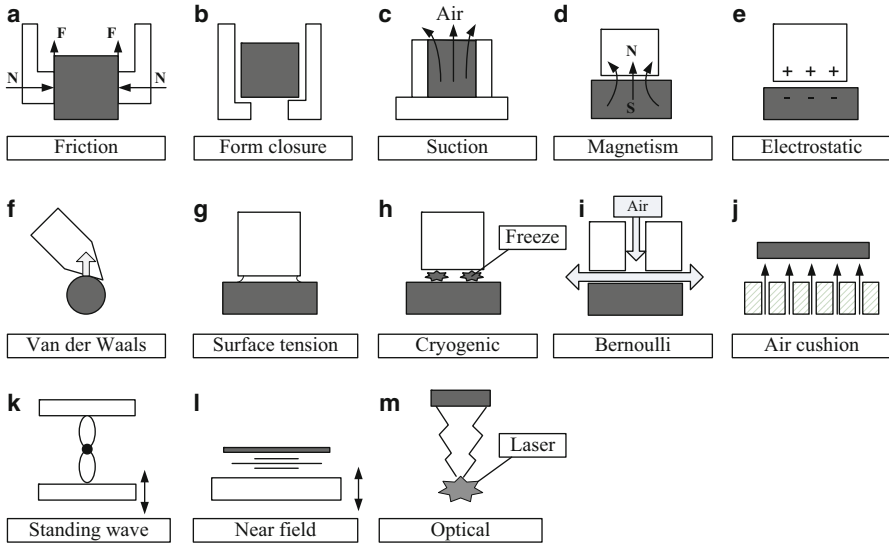
Fig. 24.1 Classification of micro-manipulating principles

total force varies. Adhesive forces become dominating while inertial forces become insignificant when an object becomes small at the micrometer level. Adhesive forces exist at the interface of a micro-object and the gripper as well. Adhesive forces are undesirable which be minimized in the operation. The gripper and the micro-object are usually interacted at contact surfaces; while not all of gripping forces come directly from the physical contacts of gripper and object.

Based on the literature review, Fig. 24.1 provides an overview of the classification of micro-handling principles; they have been classified into *friction-based gripping* [8], *form closed gripping* [9], *suction based gripping* [10], and *magnetic gripping* [11]. The aforementioned principles are applicable to macro-objects as well; the rest of principles in Fig. 24.1, in particular, the following principles, are especially applied to micro-objects.

- **Electrostatic force**. The force is produced by the different charge between gripper and micro-object [12, 13]. For a micro-object with a low weight, the capillary force and the surface tension of a liquid between gripper and object can be sufficient to hold the object in place [14].
- **Van der Waals force**. It refers to the atomic attraction between two objects. Even though such Van der Waals force is difficult to be controlled, researchers were investigating the feasibility of using it as a compensating force in gripping [15].
- **Force by changing phase**. In cryogenic gripping, a small amount of liquid is frozen at the interface of the gripper and the object, the adhesive property of ice produces the required force. To release the object, the frozen material is broken, molten, or evaporated.

- **Ultrasonic wave pressure**. Ultrasonic waves can also be used to lift up on object. Since these forces are small, only an object with very lightweight can be handled in this way. A focused light source, i.e., a laser source, can produce a pressure that is enough to lift up a micro-object. To compensate for the weight caused by the object mass, the operation might be taken place in a liquid [16]. The Bernoulli force helps to carry micro-objects.
- **Force by air pressure**. Airflow causes the pressure difference in the air channel. If the air passes through the gap of the surfaces of the gripper and the object, it generates a gripping force to close the surface gap [17].

3   Types of Grippers

In this section, different manipulating principles are applied to develop grippers for various purposes.

3.1   Grippers with Miniaturized Tweezers

This type of gripper consists of two or more miniaturized tweezers, the gripping surfaces on the tweezers make physical contacts with the object, and the friction force at the contacts holds a micro-object for gripping. To illustrate its working principle, a microgripper is designed to have two tweezers actuated by piezoelectric bimorphs. The whole gripper is packaged as the size of an electronic chip, which can be a perfect example of "plug and use" proposed by Laboratoire d'Automatique de Besan [18]. The tweezers were manufactured by the process so-called Lithographie Galvanoformung Abformung (LIGA). The main parameters of such a gripper are summarized in Table 24.1, where the gripping amplitude refers to the amplitude of the motion in the plane of two tweezers, and the insertion amplitude refers to the amplitude of the motion perpendicular to this motion plane.

Table 24.1 Properties of gripper with two tweezers

Gripper	Parameter	Value	Unit
Actuation	Piezoelectric bimorphs		
Tweezers	Length	25	mm
	Material	Nickel	
Gripping	Gripping amplitude	320	10^{-6} m
	Insertion amplitude	200	10^{-6} m
	Gripping force	80	10^{-3} N
	Insertion force	30	10^{-3} N
Dimension	Length	1	10^{-3} m
	Height	0.3	10^{-3} m

3.2 Form Closure Grippers

This type of gripper is suitable to handle sensitive elements such as microbes [9]. The manipulator creates an opening for the object to enter, approaches the microbe, closes the opening, and then captures the microbe eventually.

3.3 Vacuum Grippers

In a vacuum gripper, the pressure difference of the ambient atmosphere from a low-pressure volume inside the gripper generates a force large enough to pick up micro-objects. Such types of tools have been widespread in industry [19, 20]. Due to the adhesive forces, the requirements for picking and placing operations are opposite: the picking operation requires a large tool diameter to generate a large force; on the other hand, the tool diameter is supposed to be small to overcome adhesive force. To make a trade-off, the size of the pipette tip must be optimized for a specific size and weight of micro-object.

For example, when handling an object with the overall size of 80–150 μm, the optimized tip size ranges from 25 to 50 μm. It has been found that the optimized tip size is about 25–50 % of the object size. For a glass pipette for picking and placing operation on metallic and nonmetallic particles with the size of 50–300 μm, the rate of successful operations is around 75 %. The vacuum can be functioned by a pump with 6-voltage for 6-bar pressure supply; as a reference, the maximum output vacuum is -0.86 bar and the maximum output pressure is 6 bars. In some literatures, a vacuum gripper provides both of the suction force and a repulsive squeeze film to stop the object touch the gripper. If there is no repulsive squeeze film, a high contact force at the picking phase might lead to the formation of the cracks at the contact surface.

3.4 Magnetic Grippers

A magnetic gripper performs the manipulation by the magnetic levitation; the force comes from the magnetic field generated by magnets. Three different types of magnets can be used, i.e., *permanent magnets*, *electromagnets*, or *superconducting magnets*. A magnetic gripper has the specific requirement of materials to be manipulated; for example, the use of electromagnetic levitation is limited to materials with high electrical conductivity and to low-temperature applications [21, 22].

The magnetic levitation can be classified into two types. The first one refers to the electromagnetic system (EMS) and the second one is called electrodynamic system (EDS). In the electromagnetic system, an attractive force is generated between

normal electromagnets and a ferromagnetic guide. The equilibrium position is unstable; therefore, the closed-loop control be applied to the system to ensure the stability of manipulation [23].

Electrodynamic levitation is based on the induction of eddy currents in conducting materials. These eddy currents can be induced by a time-varying magnetic field. One type of electrodynamic systems is based on the force acting between the magnetic field generated by superconducting magnets and the stationary coils located in the guideway. The variation of the field is induced by the relative motion between the superconducting magnets and the stationary coils. The second type of electrodynamic systems is based on the force generated by a time-varying current inducing magnetic field variations. The forces from such an electrodynamic systems are repulsive, and the corresponding levitation mechanism is passively stable. Recently, the magnetic levitation has also been used to develop new force sensors for micro-assembly with the stiffness of around 0.02 N m^{-1} [24].

3.5   Electrostatic Grippers

A spherical particle near a substrate plate can be handled (such as catching, moving, and placing operations) by electrostatic effects. In such cases, all of the objects, including grippers, spherical objects, and substrates, have to be conductive. The gripping principle is to use the adhesive force to grasp objects and to impose a detachment voltage to release objects. The adhesion and the electrostatic force can be computed based on JKR model [25]. The electrostatic forces can be affected by a number of factors such as surface roughness, existence of nonconductive materials, electric discharge, tunneling current, and atmosphere humidity. Since adhesive forces affect the execution of the releasing operation, the reliability of this type of grippers becomes questionable. As far as the electric levitation is concerned, applicable materials include conductive, semiconductors, and dielectric materials. These materials can generally be classified into the following two catalogues:

(a) **Electrostatic materials**. A micro-object with electrostatic materials can be levitated by the static electrical fields induced by the polarization of object. Such functional mechanism is only suitable for polar liquids under a low-temperature field. At a high temperature, static charges cannot be sustained since they are gradually degenerated over time. On the other hand, the magnitude of the force is large to levitate an object with a relatively large size [26].

(b) **Electrodynamic materials**. For a micro-object with electrodynamic materials, it can be charged and held as a stationary state by the combination of a static (DC) and oscillating (AC) electric fields. This technique is able to lift and levitate micro-objects. However, the size of object to be levitated is relatively small, and yet, the stability of the manipulated object is poor [27].

Generally, electrostatic grippers can be applied to manipulate micro-objects with conductive materials with a satisfactory result. If a micro-object is made of

nonconductive materials, it leads to the poor accuracy of positioning and gripping forces.

3.6   Van der Waals Force Grippers

A Van der Waals force gripper manipulates a micro-object based on the atomic attraction of particles. One important consideration is to quantify some critical factors which affect the Van der Waals force of two objects. For example, increasing surface roughness increases the mean separation distance between two interacting surfaces; thus reduces the attraction force since such a force is the function of the surface distance. The larger the distance, the weaker the mutual attraction force.

3.7   Tension Grippers

A tension gripper relies on the tension forces between two surfaces to catch micro-objects by the gripping tool [28–30]. The accuracy of this gripper type can be affected by some factors such as the geometries of the gripper or micro-object and the size difference of the gripper and micro-object. The so-called centering effect occurs where the manipulated object tends to be gripped along a specific direction determined by the geometric shape of the gripper. The tension force relates to the adhesive type and volume, contact surfaces on gripper, the distance of surfaces to be contacted, and materials of the micro-object. A tension gripper seems technically feasible, but the main challenge is how to miniature a tension gripper.

3.8   Ice Grippers

An ice gripper was proposed and developed under the Eureka project by the Swiss Centre for Electronics and Microtechnology (CSEM), AP Technologies, and Sysmelec [31]. An ice gripper takes advantage of the adhesive properties of ice to pick up micro-objects. To grasp an object, a drop of water is sprayed on the surface of object, the gripper then moves to the object until it touches water. As soon as the contact is made between the gripper and the object via water, water is frozen to stick the object to the gripper, and the object can be manipulated at will. Generally, the gripping force by an ice gripper is 20–100 times stronger than the gripping force by a vacuum gripper.

To release the object, the tip of the gripper is heated up to the temperature where the phase change of ice occurs at the contacted surface. In their feasibility study, the ice gripper was able to handle objects with an overall size from 0.1 to 5 mm, the manipulating accuracy reached 1 μm, and the manipulating frequency was 1000

cycles per hour. Some identified advantages of an ice gripper are (1) high adhesion forces, (2) no damage on object surface during manipulation, (3) no sensitivity to materials to be operation, and (4) short picking and placing times [32]. The gripping principle of an ice gripper can also be referred as *phase changing gripping*.

3.9 Aerodynamic Grippers

In an aerodynamic gripper, levitating force comes from a flow of gas or liquid. Aerodynamic grippers can be classified based on the direction of flow with respect to the gripping force into *Bernoulli levitation* or *air cushion*.

(a) **Bernoulli levitation**. In Bernoulli levitation, the micro-object is placed below the manipulator, and the air flows downwardly. Due to the pressure difference, the air is able to flow radially between the gripper and the object. An increase of velocity induces a decrease of the dynamic pressure, this leads to an upward attracting force applied on the object. If the clearance gap between the gripper (more specifically, the nozzle head on the gripper) and the object is small, an increase of the flow velocity along radical directions induced by the reduction of cross-section area leads to a decrease of dynamic pressure by Bernoulli effect; this also results in an upward attracting force on the object. One should pay a special attention on the gap: when the gap becomes too large, repelling forces appear and the object can be blown away. If the gap is too small, the airflow cannot be formed effectively to product enough attracting force for object, and the object just keeps still under the gripper.

(b) **Air cushion levitation**. In air cushion levitation, the object is placed above the manipulator. The pressurized air flows upwardly through a number of air channels that are made all over the gripper; it produces a repulsive levitation force that overcomes the weight of the object [33]. The orientation of an aerodynamic force falls into two categories. The first one is the passive orientation method; wrongly oriented objects can be rejected by the means of aerodynamic baffles. The second one is the active orientation method; it requires that the object moves actively into the desired position for the manipulation.

3.10 Ultrasonic Grippers

Two configurations of ultrasonic grippers can be found in the literature: *standing wave levitation* and *near-field levitation* [34]. The first configuration is appropriate for micro-objects with a small size; they can be levitated in the pressure node with an acoustical standing wave between a vibrating plate and a reflector. The second configuration differs in sense that the reflector is moved and the levitated object serves for the purpose of reflector.

(a) **Standing wave levitation**. In the literatures, acoustic standing wave levitators are often referred to open single-axis acoustic positioners. Such grippers offer forces which are capable of carrying the particles with a weight in the order of a few grams; they can be applied in a gaseous atmosphere and under terrestrial conditions. The capacity of a standing wave levitation can be enhanced if it is applied under microgravity conditions or in a liquid environment.

To generate a standing wave, the vibrator is placed at a fixed distance from the reflector; the ideal distance is a multiple of half the wavelength. Under the microgravity condition, a micro-object would be positioned exactly in the pressure node, i.e., the point with zero acoustic pressure. Since the acoustic velocity and pressure are offset by 90° in order to satisfy the Bernoulli equation, a pressure node may also be referred as the velocity loop (point of maximum velocity) or velocity anti-node [35].

(b) **Near-field levitation**. In a near-field levitation, the object to be manipulated is placed above the manipulator surface. The separation of two surfaces can be by a high intensity vibrator or high intensity ultrasound for noncontact handling. In manipulation, the object is lifted through the direct radiation of a high intensity ultrasonic transducer underneath the near field. The gas film for load carrying is created by the rapid vibration of the manipulator. Radiation sources can be an in-phase longitudinal mode such as the piston motion or a flexural vibration mode [36]. The height of achievable levitations is relatively small; it is determined by the intensity of the deployed sound transducer. As opposite to the standing wave levitation, much large forces can be applied to objects. Theoretically, any weight can be levitated if the separating distance between the surfaces of object and the vibrating plate is small enough. A vibrator can be set up in an oblique position; this changes the levitation force and creates a potential well so that the object can return its home position automatically due to its gravity. The shape and mode of a vibrator can be tuned to induce centripetal viscous forces naturally to work against lateral sliding of the lifted object.

Acoustic radiation pressure is a topic of nonlinear acoustics which was discovered a whole century after the light radiation pressure in optics. This is mainly due to the nonlinear nature of the acoustic radiation pressure. Early researchers computing the sound wave generated in their levitator from hydrodynamic theory to give the time-averaged mean square pressure and mean square velocity of the vibrating air at each point in space. They inserted these values into a general expression for the time-averaged potential of the acoustic radiation force acting on a small rigid sphere. This expression was first derived by Lev Gor'kov [37] in 1962. The acoustic potential calculated by Xie and Wei [38] is shown as a contour plot in Fig. 24.2. This configuration corresponds to the resonant state in which the distance between emitter and reflector is approximately two sound wavelengths, which at 16.7 kHz is 20.3 mm for air at room temperature.

When working in the region near the sound source, the complex acoustic diffraction disturbs the wave field. Chu and Apfel [39] summarizes the confusion

Fig. 24.2 Acoustic potential contours in standing wave levitation [38]
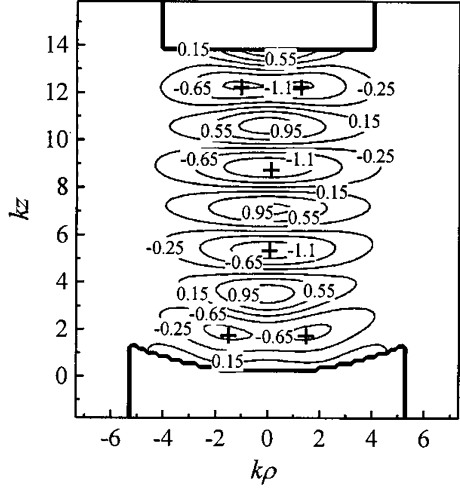
generated by acoustic radiation pressure. He calculates the Rayleigh radiation pressure successively on a plane target perfectly absorbing, partially reflecting and perfectly reflecting; he also compares Rayleigh and Langevin radiation pressures. Wiesendanger [40] first explains the basic working principle of squeeze film with very simple models. He considers infinite plates and explains how the non-linearity of the poly tropic transformation $p \cdot V^n = cst$ leads to a levitation pressure between a vibrating and a free plate. Figure 24.3 illustrated the nonlinear polytrophic transformation, the fourth quadrant shows the harmonic distance oscillation between the plates. The nonharmonic pressure oscillation is derived in the quadrant two. The mean pressure p can be found in the nonlinear squeeze film model by Li et al. [36]. In this model, p was determined by gas inertia, edge effect, and the flexural surface vibration based on the momentum equation of Newtonian fluids as,

$$\rho \left(\frac{\partial v_r}{\partial t} + v_r \frac{\partial v_r}{\partial r} + v_z \frac{\partial v_r}{\partial z} \right) = -\frac{\partial p}{\partial r} + \frac{\partial \tau_{rz}}{\partial z} \tag{24.1}$$

where $\tau_{rz} = \mu(\partial v_r/\partial z)$ is the shear stress; p, ρ, and μ are gas pressure, density, and viscosity, respectively; v_r and v_z are the velocities component along r and z. The pressure drop at the edge was calculated according to the conservation of mechanical energy in fluid. The analytical solution was derived by the mean volume average across the film thickness. According to the presented model in [36], exemplifying non-linearity of the polytropic transformation is illustrated in Fig. 24.3.
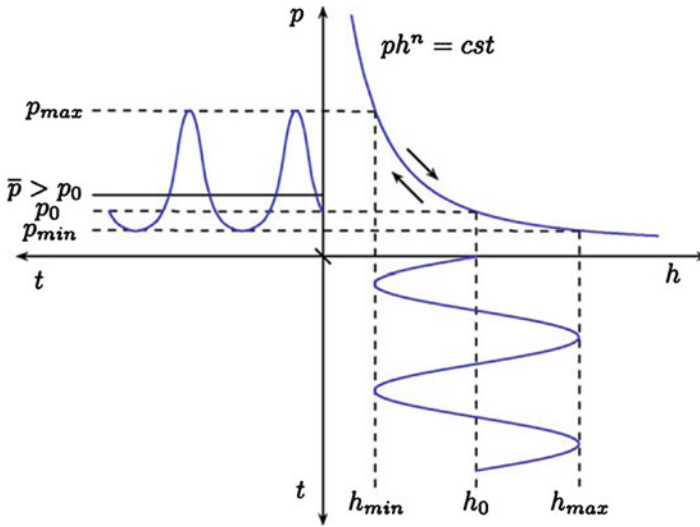
Fig. 24.3 Non-linearity of the polytropic transformation [40]

3.11 Optical or Laser Grippers

Optical levitation was first discovered by Ashkin [41]. Micro-objects such as particles can be freely suspended or accelerated by the forces caused by the radiation pressure from visible laser light. Stable potential wells exist where particles can be trapped. However, the main problem is obscuring effect of the thermal force caused by temperature gradients. Temperature gradients can be induced in the medium which surrounds the object. The manipulating forces in an optical gripper are called as radiometric forces; these forces are caused by the photophoresis effect as far as the light is concerned. A thermal force might make the levitated position of object unstable, but this can be addressed by suspending relatively transparent particles in transparent surrounding media. Transparent particles can be made of water, oil, or air.

Due to both of the beam reflection and refraction, the particle ensures two types of forces, i.e., *scattering force* and *gradient force*. The so-called scattering force is proportional to the optical intensity; it points to the direction of the incident beam. The gradient force is proportional to the intensity gradient. When the refractive index of the scatterer is greater than that of the surrounding medium, the gradient force points to the positive direction of the intensity gradient [42]. If the relative magnitude of the indices is reversed, the sign of the radial force is changed and the sphere is pushed out of the beam [43].

Using different scattering and gradient forces, many varieties of optical traps have been proposed. For example, if the single laser beam is weakly converged, the particle will be trapped to the beam center but propelled along the beam. The

stability along the beam axis relies on the balance of the scattering force and the gravity. A laser beam trap can also be composed of a highly convergent beam, which is commonly called as an optical tweezer.

The axial gradient force dominates the axial stability and the particles are trapped near the beam focus. An alternative trapping method is to use two or more laser beams to generate one stable equilibrium point, which is located below the beam-crossing point.

4 Summary

Micromanipulation is one of critical research subjects in making the products of microelectromechanical systems (MEMS). The manipulation mechanisms and tools for micro-objects are significantly different from those for macro-objects. In this chapter, we focus on the characteristics of the micromanipulation in terms of the types and principles of gripping forces. We find that different gripping principles apply to different micro-objects; there is no universal principle, which can be applicable to any micro-object in the size range between 10 μm and 1 mm.

For other manipulation principles especially for micro-objects, applicable principles are those which are able to produce van der Waals, electrostatic, surface tension forces, and the forces from pressure difference, buoyancy, and phase changing. The type of vacuum grippers is an interesting alternative to the downscaled mechanical gripper; it allows handling micro-objects placed on a flat plane. A two-finger gripper is not able to pick an object from a plane if the height of the object is insignificant in comparison with the finger size. The type of surface tension grippers is also very promising in sense that the capillary force is the ideal approach to deal with micro-objects; the overall size of micro-objects varies from the nanoscale (capillary condensation) up to the submillimetric scale. Electrostatic forces can be one good candidate as handling tools of micro-objects. However, the grippers with electrostatic forces meet the challenge from electromagnetic perturbations. Such perturbations exist widely on micro-components. Another limitation of the gripper using electrostatic force is the possibility of electrostatic forces in disturbing the manipulation if the gripper is actuated based on the electrostatic principle. Finally, despite of its long responsive time, thermal actuation is currently envisaged as a favorable actuation in a micro-manipulator, it uses a high electric field to produce the force.

We argue that one of the main issues the micromanipulation undergoes is how to deal with the adhesion forces occurring to contact surfaces. We propose several main strategies to address this issue. For releasing operations, the proposed strategies are to reduce or completely remove adhesion forces; for picking operations, we take advantages of adhesion forces to grasp objects. Adhesion forces can be completely avoided by sustaining the separation of surfaces; a number of manipulating principles can be applied without the surface contacts of the gripper and the object. Note that the advantages of noncontact handling are applicable to the manipulation of macro-objects as well.

References

1. Park JK, Ro PI (2013) Noncontact manipulation of light objects based on parameter modulations of acoustic pressure nodes. J Vib Acoust 135:031011 (1–7)
2. Whitworth G, Coakley W (1992) Particle column formation in a stationary ultrasonic field. J Acoust Soc Am 91:79–85
3. Whitworth G, Grundy M, Coakley W (1991) Transport and harvesting of suspended particles using modulated ultrasound. Ultrasonics 29(6):439–444
4. Woodside SM, Bowen BD, Piret JM (1997) Measurement of ultrasonic forces for particle–liquid separations. AIChE J 43(7):1727–1736
5. Tuziuti T, Kozuka T, Mitome H (1999) Measurement of distribution of acoustic radiation force perpendicular to sound beam axis. Jpn J Appl Phys 38:3297–3301
6. Spengler J, Coakley W, Christensen K (2003) Microstreaming effects on particle concentration in an ultrasonic standing wave. AIChE J 49(11):2773–2782
7. Perales F, Gonzalez I (2005) On the forces acting in micromanipulation of particles at low frequencies. In: Proceedings of the IEEE international ultrasonics symposium, Rotterdam, The Netherlands, 18–21 Sept, vol 4, pp 2108–2111
8. Kohl M, Krevet B, Just E (2002) SMA microgripper system. Sens Actuators A 97–98:646–52
9. Ok J, Chu M, Kim C-J (1999) Pneumatically driven microcage for micro-objects in biological liquid. In: IEEE MEMS, pp 459–563
10. Petrovic D, Popovic G, Chatzitheodoridis E, Del Medico O, Almansa A, Sumecz F, Brenner W, Detter H (2002) Gripping tools for handling and assembly of microcomponents. In: Proceedings of the 23rd international conference on microelectronics (Miel 2002), 12–15 May, NIS, Yugoslavia, vol 1, pp 247–250
11. Zesch W, Brunner M, Weber A (1997) Vacuum tool for handling micro-objects with a NanoRobot. In: IEEE international conference on robotics and automation, 20–25 Apr, vol 2, pp 1761–1766
12. Enikov ET, Lazarov KV (2001) Optically transparent gripper for microassembly. Proc SPIE 4568:40–9
13. Tsuchiya K, Murakami A, Fortmann G, Nakao M, Hatamura Y (1999) Micro assembly and micro bonding in nano manufacturing world. Proc SPIE 3834:132–40
14. Bark C, Binnenböse T, Vögele G, Weisener T, Widmann M (1998) Gripping with low viscosity fluids. In: IEEE international workshop on MEMS, pp 301–305
15. Arai F, Fukuda T (1997) A new pick up and release method by heating for micromanipulation. In: IEEE MEMS, pp 383–388
16. Bancel PA, Cajipe VB, Rodier F, Witz J (1998) Laser seeding for biomolecular crystallization. J Cryst Growth 191:537–44
17. Grutzeck H, Kiesewetter L (2002) Downscaling of grippers for micro assembly. Microsyst Technol 8:27–31
18. Agnus J (2003) Contribution à la micromanipulation: Etude, Réalisation, Caractérisation et Commande d'une Micropince Piézoélectrique. PhD thesis, Laboratoire d'Automatique de Besançon (UMR CNRS 6596)
19. Danuser G, Pappas I, Vögeli B, Zesch W, Dual J (1998) Manipulation of microscopic objects with nanometer precision: potentials and limitations in nano robot design. Int J Rob Res (submitted)
20. Zesch W, Brunner M, Weber A (1997) Vacuum tool for handling microobjects with a nanorobot. In: Proceedings of the international conference on robotics and automation, pp 1761–1766
21. Kang BJ, Hung LS, Kuo SK, Lin SC, Liaw CM (2003) H∞ 2DOF control for the motion of a magnetic suspension positioning stage driven by inverter-fed linear motor. Mechatronics 13(7):677–696
22. Motokawa M, Mogi I, Tagami M, Hamai M, Watanabea K, Awaji S (1998) Magnetic levitation experiments in Tohoku University. Physica B 256:618–620

23. Bona B, Brusa E, Carabelli S, Chiaberge M, Delprete C, Genta G et al (1997) Review article: the mechatronics laboratory at politecnico di torino. Mechatronics 7(5):413–427

24. Boukallel M, Piat E, Abadie J (2003) Passive diamagnetic levitation: theoretical foundations and application to the design of a micro-nano force sensor. In: Proceedings of the 2003 IEEE international conference on intelligent robots and systems, Las Vegas, Nevada, pp 1062–1067

25. Johnson KL, Kendall K, Roberts AD (1971) Surface energy and the contact of elastic solids. Proc R Soc Lond Ser A 324:301

26. Fantoni G (2003) Assembly of mini and microparts: development of an electrostatic feeder. In: Proceedings of the 6th A.I.Te.M. international conference

27. Bar-Ziv A, Sarofim AF (1991) The electrodynamic chamber: a tool for studying high temperature kinetics involving liquid and solid particles. Prog Energy Combust Sci 17:1–65

28. Bark C, Binnenboese T (1998) Gripping with low viscosity fluids. In: Proceedings of IEEE international workshop on MEMS, Heidelberg, pp 301–305

29. Bark K-B (1999) Adhäsives Greifen von kleinen Teilen mittels niedrigviskoser Flüssigkeiten. Springer, Berlin

30. Biganzoli F, Fassi I, Pagano C (2005) Development of a gripping system based on capillary force. In: Proceedings of ISATP05, Montreal, Canada, pp 36–40

31. El-Khoury M (1998) Ice gripper handles micro-sized components. Design News. www.designnews.com

32. Stefan J, Seliger G (1999) Handling with ice—the cryo-gripper, a new approach. Assem Autom 19(4):332–337

33. Gengenbach U, Boole J (2000) Electrostatic feeder for contactless transport of miniature and microparts. In: Proceedings of SPIE conference on microrobotics and microassembly II, pp 75–81

34. Höppner J (2002) Verfahren zur berührungslosen handhabung mittels leistungsstarker schall-wandler. PhD thesis, Technische Universität München

35. Daidžić N (1995) Nonlinear droplet oscillations and evaporation in an ultrasonic levitator. PhD thesis, Universität Erlangen-Nürnberg

36. Li J, Cao W, Liu P, Ding H (2010) Influence of gas inertia and edge effect on squeeze film in near field acoustic levitation. Appl Phys Lett 96:243507

37. Brandt EH (2001) Suspended by sound. Nature 413:474–475

38. Xie W, Wei B (2001) Parametric study of single-axis acoustic levitation. Appl Phys Lett 79:881

39. Chu BT, Apfel RE (1982) Acoustic radiation pressure produced by a beam of sound. J Acoust Soc Am 72(6):1673–1687

40. Wiesendanger M (2001) Squeeze film air bearings using piezoelectric bending elements. PhD thesis, Ecole polytechnique fédérale de Lausanne

41. Ashkin A (1970) Acceleration and trapping of particles by radiation pressure. Phys Rev Lett 24(4):156–159

42. Ashkin A, Dziedzic JM, Bjorkholm JE, Chu S (1986) Observation of a single beam gradient force optical trap for dielectric particles. Opt Lett 11(5):288–290

43. Bancel P, Cajipe V, Rodier F (1999) Manipulating crystals with light. J Cryst Growth 196:685–690

Chapter 25
Inertial Microfluidics: Mechanisms and Applications

Jun Zhang, Weihua Li, and Gursel Alici

Abstract In microfluidics, the typical sample volume is in the order of nL, which is incompatible with the common biosample volume in biochemistry and clinical diagnostics (usually ranging from 1 μL to approximately 1 mL). The recently emerged inertial microfluidic technology offers the possibility to process large volume ($\sim$mL) of biosample by well-defined micro-structures. In contrast to conventional microfluidic technologies, where fluid inertia is negligible and flow remains almost within Stokes flow region with very low Reynolds number $Re \ll 1$ ($Re = \rho_f \overline{U}_f H / \mu$, where ρ_f, $\bar{U}_f$ and μ are fluid density, average velocity and dynamic viscosity, respectively, and H is channel hydraulic diameter), inertial microfluidic devices work within an intermediate Reynolds number range ($\sim 1 < Re < \sim 100$) between Stokes and turbulent regimes. In this intermediate range, both inertia and fluid viscosity are finite, and several intriguing effects appear and form the basis of inertial microfluidics, including (i) inertial migration and (ii) secondary flow. Due to the superior features of high-throughput, simplicity, precise manipulation and low-cost, inertial microfluidics has attracted significant attention from the microfluidic community. Meanwhile, a number of channel designs that focus, concentrate and separate particles and fluids have been explored and demonstrated. In this chapter, we discuss this fascinating technology from three crucial aspects: (1) fundamental mechanism, (2) microchannel designs and (3) applications. From this chapter, we hope that readers can have a clear understanding on the concept of inertial microfluidics, its working mechanism and potential applications.

J. Zhang • W. Li (✉)
School of Mechanical, Materials and Mechatronic Engineering, University of Wollongong, Wollongong, NSW 2522, Australia
e-mail: weihuali@uow.edu.au

G. Alici
School of Mechanical, Materials and Mechatronic Engineering, University of Wollongong, Wollongong, NSW 2522, Australia

ARC Center of Excellence for Electromaterials Science, University of Wollongong, Wollongong, NSW 2522, Australia

Notations

a	Particle diameter
AR	Aspect ratio of channel ($=h/w$)
D_{max}	Rotational diameter of non-spherical particle
De	Dean number
f_{drag}	Viscous drag coefficient
f_L	Coefficient of net inertial lift force
F_D	Secondary flow drag or Dean drag
F_{drag}	Viscous drag force
F_L	Net inertial lift force
F_{LR}	Magnus force or rotation-induced lift force
F_{LS}	Shear gradient lift force
F_{LW}	Wall lift force
F_S	Saffman force or slip-shear-induced lift force
h	Channel height
H	Channel hydraulic diameter
L_{min}	Minimum channel length for particles to migrate to the inertial equilibrium position
R	Radius of curvature of curving channel
Re	Reynolds number
Re'	Particle Reynolds number based on relative velocity of fluid and particle
R_f	Ratio of inertial lift force to Dean drag
R_p	Particle Reynolds number based on the size ratio of particle to channel
S	Cross-sectional area of particle
U_D	Secondary flow velocity or Dean flow velocity
U_f	Fluid velocity
U_p	Particle velocity
v_t	Relative velocity of fluid to particle
V_t	Relative velocity of particle to fluid ($=-v_t$)
w	Channel width
x	Lateral position of particle
γ	Fluid shear rate
μ	Dynamic viscosity
ρ_f	Fluid density
υ	Kinetic viscosity
Ω_p	Angular velocity of particle

1   Introduction

The inertial migration (or the tubular pinch effect) was first observed by Segre and Silberberg in 1961 [1, 2]. In their experiments, spherical particles (∼mm in diameter) migrated to an annulus located about 0.6 times of pipe radius between centreline and pipe wall in a cylindrical pipe (∼cm in diameter). After that, scientists have been spending decades to understand the underlying physics of this intriguing phenomenon through experimental studies, theoretical analyses and numerical simulations [3–10]. Although the exact mechanism is still unclear, it has been widely accepted that the two opposing lateral lift forces are responsible for the phenomenon: (1) the wall lift force, F_{LW}, a result of the flow field interaction between the suspended particles and the adjacent walls, which repels the particles away from the wall; (2) the shear gradient lift force F_{LS}, due to the curvature of the fluid velocity profile and its interaction with the particles, which directs particles away from channel centre. The final balance between these inertial lift forces results in a set of dynamic equilibrium positions within the cross section of channels dependent on the cross-sectional shape [11–13], Reynolds number [14], particle shape [15], deformability [16], etc. It should be noted that there are actually two additional lateral forces acting on the particles in a wall-bounded Poiseuille flow: (1) rotation-induced lift force (or Magnus force); and (2) slip-shear-induced lift force (or Saffman force). However, their magnitudes are generally at least one order less than that of the shear gradient lift force [17, 18], which are, therefore, negligible in the common inertial microfluidic analysis.

Since the first observation of inertial migration, all the investigation was focused on the understanding of its fundamental mechanism. Until the recent emergence of microfluidics, where the size of patterned micro-channel is comparable to that of the suspended particles (so that inertial migration can be more obvious within a short channel length), it has finally found its application in biomedicine and industry. Then, inertial microfluidics, an emerging field in microfluidics, that utilizes inertial migration and other inertial effects (secondary flow and centrifugal force) to realize the manipulation of fluid and suspended particles, was defined by its characteristics of finite inertial flow [19–23].

In contrast to the common sense that, in microfluidics, fluid inertia is negligible (Stokes flow regime, $Re \rightarrow 0$), inertial microfluidics works in an intermediate range ($\sim 1 < Re < \sim 100$) between Stokes regime ($Re \ll 1$) and turbulent regime ($Re \sim 2000$), where both inertia and viscosity of the fluid are finite [18]. Besides the intriguing phenomenon of inertial migration, there are additional inertial effects due to the finite inertia of fluid (secondary flow) and particle (centrifugal force). And these inertial effects together constitute inertial microfluidics.

The particle centrifugal force, which is proportional to the density difference between particle and medium [24–26], is normally negligible in inertial microfluidics for biomedical applications, where the density of cells or particles is very close to that of the medium [27, 28]. Meanwhile, secondary flow usually appears in a curved channel [29] or a straight channel with disturbance structures [30–32]. In a curved channel, secondary flow is induced by a pressure gradient in the

radial direction because of fluid momentum mismatch in the centre and near-wall region within the cross section [20]. The fluid elements near the channel centreline have a higher momentum than that near the wall, therefore flow outwards, and drive relatively stagnant fluid elements near the channel wall inwards along the circumference, forming two counter-rotating streams, well known as Dean vortices [29]. Introducing secondary flow in inertial focusing brings several benefits. For example, Dean flow can modify the inertial equilibrium positions by imposing an additional viscous drag force on particles perpendicular to the main stream. Also, size-dependent differential focusing of particles dependent on the ratio of inertial lift force and secondary flow drag (F_L/F_D) promises a complete particle separation. In addition, Dean flow could reduce the channel length/footprint, due to the mixing effects of secondary flow, assisting particles to reach stable equilibrium positions more quickly [20].

Inertial microfluidics has attracted a great attention during the last decade due to its advantages of precise manipulation, simple structure and high-throughput. In contrast to numerous active microfluidic manipulation technologies [33–40] where one or more external force fields are supplied to control the motion of target particles or cells, inertial microfluidics is a passive manipulation technology, and it manipulates and separates bioparticles by the intrinsic hydrodynamic force [41]. Therefore, the operation of inertial microfluidics is very simple and robust, and the cost of inertial microfluidic device is relatively low. Current investigation in the area of inertial microfluidics is mainly focused on three aspects: (1) designing novel micro-channel structures, (2) uncovering the fundamental mechanisms of inertial focusing within them and (3) exploring and expanding their applications in biomedicine and industry. This chapter is organized as follows: first, we will discuss the fundamental theory behind the inertial migration and particle focusing in microfluidics, so that readers can have a clear understanding why the inertial focusing happens, and how its focusing pattern can be altered. Second, the current design of functional micro-channels in inertial microfluidics will be presented. After that, we will summarize the current developments of inertial microfluidic technology in biomedicine, disease diagnosis and industry. Finally, conclusions and perspectives of inertial microfluidics will be proposed.

2   Fundamental Mechanism

In inertial microfluidics, when the particle suspension is infused into a microfluidic channel, there are several forces acting on the suspended particles, and they determine particles' motion along the mainstream (axial) and lateral direction (Fig. 25.1). Along the mainstream, the viscous drag force, arising due to the velocity difference between particles and fluid, accelerates particles close to the velocity of its surrounding fluid elements. When particles are near a channel wall, the wall has an effect to slow down the motion of particles, so that particles are actually lagging the fluid flow slightly. On lateral direction that is perpendicular to the fluid mainstream, there are also a number of lateral lift forces, such as rotation-

Fig. 25.1 Mainstream and lateral forces exert on particles in a microfluidic channel
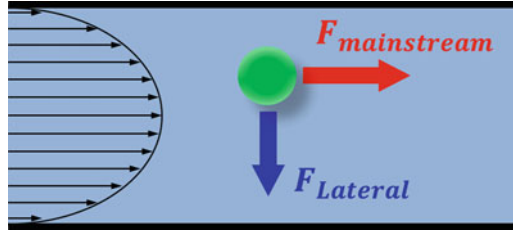
induced lift force, slip-shear-induced lift force, wall lift force, shear gradient lift force, deformability-induced lift force and secondary flow drag force [42–45]. In the real inertial microfluidic analyses and applications, not all of these forces need to be taken into account. Some of them are actually negligible, depending on the property of medium and particles and structure of channels.

2.1 Viscous Drag Force

Drag force arises when an object moves through a fluid or, relatively, when the fluid flows toward an object. The origin of the drag force lies in the need to remove the elements of fluid out of the way of a moving object. The drag force on a moving spherical particle can be expressed as:

$$F_{\mathrm{drag}} = S * f_{\mathrm{drag}} = \pi a^2 f_{\mathrm{drag}}/4 \tag{25.1}$$

where S is the cross-sectional area of the particle, and a is the diameter of the particle. f_{drag} is viscous drag coefficient which is always determined by the particle Reynolds number $Re' = v_t \rho_f a/\mu$, here v_t is the relative velocity of fluid to particle [46].

Although the viscous drag coefficient f_{drag} has different expressions according to the range of particle Reynolds number Re', the most widely used is the Stokes drag due to its simplicity, especially when relative velocity of fluid to particle v_t is very small [46].

$$f_{\mathrm{drag}} = 12\frac{\mu v_t}{a} \tag{25.2a}$$

Then,

$$F_{\mathrm{drag}} = 3\pi \mu a v_t \tag{25.2b}$$

In inertial microfluidics, viscous drag force consists of two components, respectively, along (1) mainstream direction, due to the axial velocity difference between fluid and the suspended particles; and (2) lateral direction, due to the secondary flow induced by the channel curvature, or disturbance structure.

Fig. 25.2 Rotation-induced lift force for a rigid cylinder and sphere in a uniform flow
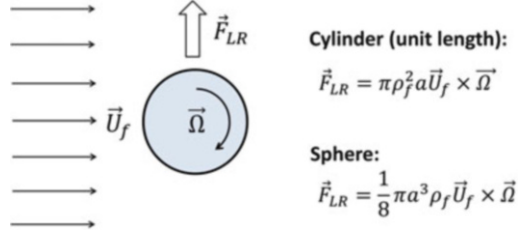
2.2 Magnus Force–Rotation-Induced Lift Force

In an in-viscous flow with uniform velocity U_f, a stationary cylinder is rotating with constant angular velocity Ω, see Fig. 25.2. Assuming no slip velocity at the surface of the cylinder, the fluid velocity at the bottom part is lower than that at the upper part. According to the Bernoulli principle, fluid pressure is higher at the lower part than that in the upper part. As a result, a lift force F_{LR} is developed to lift the cylinder. The magnitude of this lift force, per unit length of the cylinder is [42]:

$$\overrightarrow{F}_{\text{LR}} = \pi \rho_f^2 a \overrightarrow{U}_f \times \overrightarrow{\Omega} \tag{25.3}$$

Similarly for a rigid sphere rotating in a fluid, a lateral lift force appears due to the transverse pressure difference, and this force is called "Magnus force" [47]:

$$\overrightarrow{F}_{\text{LR}} = \frac{1}{8} \pi a^3 \rho_f \overrightarrow{U}_f \times \overrightarrow{\Omega} \tag{25.4}$$

If the sphere is not stationary, but simultaneously translating through the fluid with a velocity $\overrightarrow{U}_p$, replacing the fluid velocity vector $\overrightarrow{U}_f$ with the relative velocity $\left(\overrightarrow{U}_f - \overrightarrow{U}_p \right)$.

The direction of Magnus force is perpendicular to the plane defined by the vectors of the relative velocity and the axis of rotation. In the case of a sphere rotating with an angular velocity, $\overrightarrow{\Omega}_p$ in a rotational flow field, the vector $\overrightarrow{\Omega} \left(= \overrightarrow{\Omega}_p - 0.5 \nabla \times \overrightarrow{U}_f \right)$ represents the relative rotation between fluid and sphere.

Magnus force can be considered as a result of pressure difference induced by the streamline asymmetry, due to the rotation of sphere. Although above particular expressions (25.3)–(25.4) are only valid for low Reynolds numbers, this kind of lift force due to the rotation of a body is also present in more strongly inertial flow [42, 48].

2.3 Saffman Force–Slip-Shear-Induced Lift Force

Using the matched asymptotic expansion method, Saffman [43] calculated the lateral lift force on a sphere in a simple unbounded shear flow. The magnitude of this force is:

$$F_S = \frac{K}{4}V_t a^2 \left(\gamma v^{-1}\right)^{1/2} \tag{25.5}$$

where K is a numerical constant ($K \sim 81.2$), γ is shear rate, v is the kinetic viscosity and V_t is the relative velocity between particle and fluid at the streamline through the centre of particle. It should be noted that this force is the only effect of shear rate, independent of rotation of particle. In the real situation of force-free spheres, a relative rotation will be induced by the fluid shear. Consequently, a Magnus-type force may appear on the spheres. In most practical flows, the effects of rotation and shear cannot be considered as independent, and a simple superposition of the two lift forces would inevitably bring errors [42]. Unless the rotation speed is much greater than the shear rate, for a freely rotating particle $\Omega_p = \frac{1}{2}\gamma$, Saffman force F_S, caused by the interaction of slip velocity (Stokeslet flow field) and shear, will generally be at least one order of magnitude larger than Magnus force F_{LR} caused by the interaction of slip and particle rotation at low Reynolds numbers [48].

The direction of Saffman force is always toward the side, where the magnitude of relative velocity V_t is maximum. If the particles are leading the flow, Saffman force directs to the stagnant wall in a simple shear flow; Fig. 25.3a; if the particles are lagging the flow, Saffman force points to the moving wall, see Fig. 25.3b. Saffman force is more relevant in the case of non-neutrally buoyant particles in a vertical flow, or additional outer force (electrical or magnetic) [18, 49] acting on the particle to lag or lead fluid flow. In these cases, the Saffman force will direct to the channel centreline when particles lag the flow, or direct toward the channel walls when particles lead the flow in Poiseuille flow.
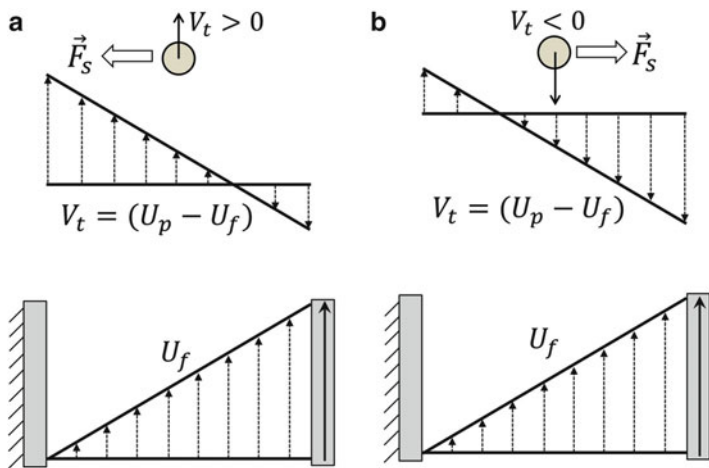
Fig. 25.3 Saffman force on a sphere in a simple shear flow. V_t is the relative velocity of particle to fluid. (**a**) $V_t > 0$ indicates that particle moves faster than the corresponding fluid element, leading the flow. (**b**) $V_t < 0$ means that particle moves slower than the fluid element, lagging the flow. The direction of lateral lift force always directs toward the side, where the magnitude of relative velocity V_t is maximum

2.4   Wall Lift Force

The existence of walls creates a velocity gradient (shear rate) of fluid and causes the rotation of spherical particles due to the fluid shear. Thus, transverse Magnus force and Saffman force may act on the particles to migrate particles laterally. In addition, the secondary part of wall effects is the variation of the flow field around particles due to the presence of the physical walls. In order to eliminate effects of shear rate to the lateral lift force, one possible scheme is to investigate the motion of dense particles near a vertical wall in a stagnant fluid. In general, the effect of wall on the motion of immersed objects is to retard the motion of the object in both parallel and perpendicular directions, as well as exerting a transverse migration motion. Two distinct interactions between an immersed object and walls exist: (i) The motion of the immersed object is primarily influenced by a single wall on one side of the object. Any other walls are too far from the immersed object to influence significantly its motion. In this case, the main effect of wall is decelerating the particle and driving the immersed object away from the wall by a wall lift force F_{LW} [50], Fig. 25.4a. It applies to the situation where the characteristic dimension of immersed particle is much less than that of the channel. (ii) The immersed object is surrounded by the boundaries (walls) of the flow domain. Thus, the boundaries decelerate significantly the motion of object and move particles to the centreline of channel. It happens if the characteristic dimension of the immersed particle is of the same order of the dimension of flow channel, e.g. small bubbles moving in capillaries [42].
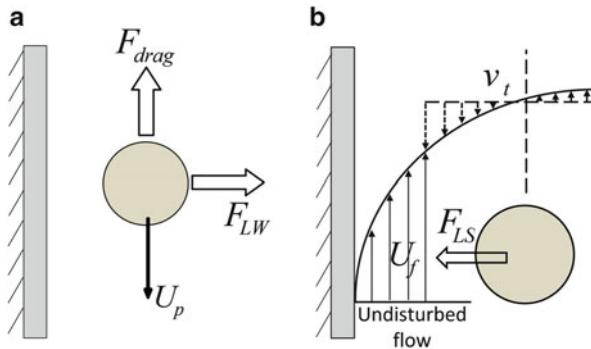
Fig. 25.4 (**a**) Boundaries retard the motion of the particle, and lateral wall lift force F_{LW} directing perpendicularly to the main flow direction repels particles away from the wall. (**b**) Shear gradient lift force F_{LS} on a particle in a Poiseuille flow. In the moving frame with the immersed particle, the relative velocity of fluid to particle v_t is larger on the wall side due to the parabolic velocity profile of fluid. Therefore, shear gradient lift force F_{LS} directs toward the wall

2.5 Shear Gradient Lift Force

As discussed above, the particle will lag the flow due to the effects of wall. If the curvature of the undisturbed fluid velocity profile is zero, it becomes a simple shear flow. Then, the pressure will be higher on the left, pushing particles to the centreline of channel, as shown in Fig. 25.3b. However, in the Poiseuille flow, the magnitude of relative velocity of fluid to particle v_t is much higher on the left side of particle than that on the right side, due to the parabolic nature of velocity profile, as shown in Fig. 25.4b. And it even overwhelms the asymmetry caused by lag velocity. Similar to Saffman force, the dissymmetry of relative velocity causes a low pressure on the left/wall side, generating a shear gradient lift force F_{LS} that's opposite to the wall lift force F_{LW}. The shear gradient lift force leads particles toward the wall until the wall lift force repels and balances it [48, 51].

2.6 Net Inertial Lift Force

For a neutrally buoyant rigid sphere flowing in a straight wall-bounded Poiseuille flow, besides a viscous drag force along the main axis, there are four lateral forces acting on the sphere as we discussed above: Magnus force due to slip-rotation, Saffman force due to slip-shear, wall lift force due to the disturbance of flow field around particles from wall, and shear gradient lift force due to the curvature of undisturbed fluid velocity profile. Among them, shear gradient lift force is at least one order of magnitude greater than the Saffman force, and three orders of magnitude greater than Magnus force [17, 44]. Therefore, both the shear gradient lift force and the wall lift force are widely accepted as the dominant effects for the inertial migration of particles [18]. The balance of the shear gradient lift force and the wall lift force creates several equilibrium positions halfway between the walls and centreline of the channel, and this theory can explain the observation of Segre and Silberberg reasonably well [1, 2], Fig. 25.5a.

Through the method of matched asymptotic expansions, Asmolov [6] derived an analytical expression of the net inertial lift force acting on a small rigid sphere ($a/H \ll 1$) in a Poiseuille flow:

$$F_L = f_L \rho_f \gamma^2 a^4 \tag{25.6}$$

The force can be further simplified as:

$$F_L = f_L \rho_f \overline{U}_f^2 a^4 / H^2 \tag{25.7}$$

The hydraulic diameter H *is* defined as $H = D$ for a circular channel (D is the diameter of circular cross section) or $H = 2wh/(w+h)$ for a rectangular channel (w and h correspond to the width and height of the rectangular cross section). The lift
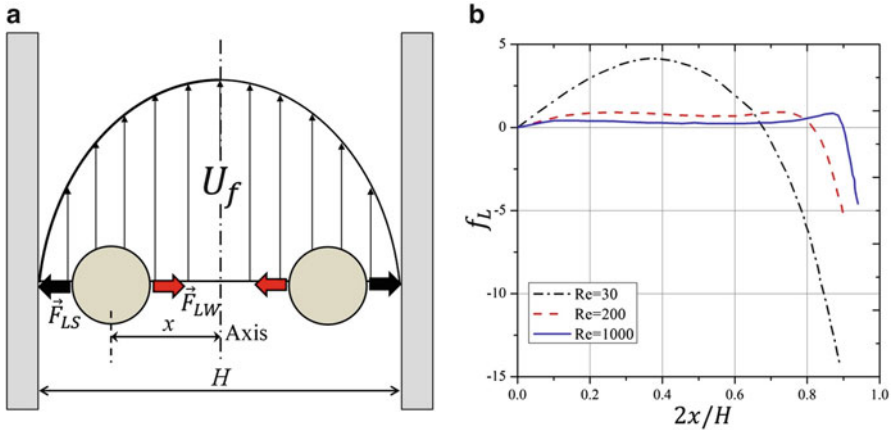
Fig. 25.5 (**a**) Balance of shear gradient lift force and wall lift force results in the inertial equilibrium positions in a Poiseuille flow. (**b**) The net lift coefficient is a function of particle lateral position x and Reynolds number Re. Reproduced from reference [67]

coefficient f_L is a function of the particle lateral position x and channel Reynolds number Re [6, 52], Fig. 25.5b. The lateral position x_{eq}, where $f_L = 0$ corresponds to the inertial equilibrium position. It should be noted that the channel centreline $x = 0$ is not a stable equilibrium position since a little deflection from centreline will never return the particles back. In addition, it has found that f_L decreases with increasing Re (or $\bar{U}_f$), suggesting that inertial lift force scales less strongly than $\bar{U}_f^2$ [12, 20].

The recent investigation suggested that finite size of particle impacts the inertial migration of particles. For a finite-size particle ($0.05 \leq a/H \leq 0.2$), particle would cause disturbance to the main channel flow. Near the channel centre, where the effects of the wall are weak, the lift force scales as $F_L \propto \rho_f \overline{U}_f^2 a^4/H^2$. While near the channel wall, where the wall effects dominate, it can be scaled as $F_L \propto \rho_f \overline{U}_f^2 a^6/H^4$ [45].

2.7 Secondary Flow Drag

If particle suspension is not flowing in a straight channel, but a channel with curvatures, e.g. spiral or serpentine, a secondary flow arises due to the velocity mismatch in the downstream direction between fluid in the central and near-wall regions. The fluid elements near channel centreline have higher inertia and would tend to flow outward around a curve due to the centrifugal effect, creating a pressure gradient in the radial direction within the channel. In a fully bounded channel, due to the centrifugal pressure gradient, relatively stagnant fluid near the walls re-circulates inward, finally forming two symmetric circulating vortices [20], called
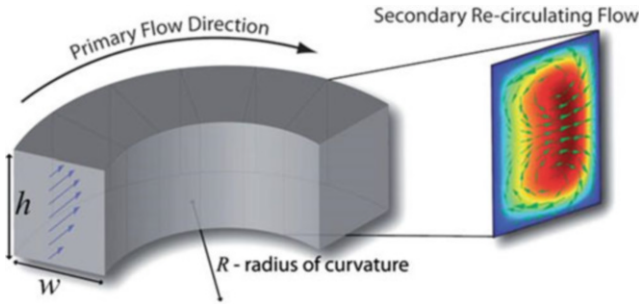
Fig. 25.6 Dean flow with two counter-rotating vortices appears in a curved channel. Reproduced from [20]

Dean vortices, as shown in Fig. 25.6. The parameters that influences the distribution and strength of the secondary flow in a curved channel includes Dean number $De\left(= (H/2R)^{1/2}Re\right)$, Reynolds number Re, aspect ratio of the channel AR $(=h/w)$ as well as the shape of cross section [29, 53].

Following Squires and Quake [54], the secondary flow velocity scales as

$$U_{\mathrm{D}} \sim De^2 \mu / (\rho H) \tag{25.8}$$

Furthermore, the magnitude of secondary flow U_{D} can be approximated as [55, 56]:

$$U_{\mathrm{D}} = 1.8 \times 10^{-4} \times De^{1.63} \tag{25.9}$$

Besides curved channels, secondary flow also appears in straight channels with disturbance structures (e.g. expansion–contraction arrays, pillar array) due to the similar mechanism. Moreover, Dean-like vortices can even be observed in specific positions of these channels [31]. Here, secondary flow drag is used to represent the viscous drag force due to the effects of Dean vortices in curved channels and other kinds of flows (Dean-like vortices) in the cross section of microchannel.

2.8 Other Effects: Particle Shape, Deformability, Density and Fluid Viscoelasticity

Although solid rigid particles can be used as a simple model in the study of hydrodynamic behaviour of particles in a microchannel, the practical bioparticles such as cells and vesicles are not rigid but deformable. The deformability will induce additional lift force on the particles. The deformability-induced lift force is perpendicular to the main streamline, which is believed to be the effects of shape-

change of particle and nonlinearities caused by the matching of velocities and stresses at the deformable particle interface [18]. In inertial microfluidics, centre-directed deformability-induced lift force shifts inertial equilibrium positions a little closer to the channel centreline than that of rigid particles [16].

Generally, for the applications of inertial microfluidics in biomedicine, the density of particle is often assumed to be equal to that of the medium, and the centrifuge force is ignored in the analysis and application. However, under certain circumstances where particle density is significantly greater than that of medium, centrifugal force on particles may need to be taken into account [44]. Besides, the shape of a particle also influence particle inertial focusing, and the rotational diameter D_{max} of a particle was found to determine the final focusing position [15]. By employing the difference in shape of particles, it has demonstrated the possibility of separating particles that have similar dimensions along one axis but that varies along another axis [57].

In addition, previous investigations were restricted within the regime of Newtonian fluid. Recently, by addition of polymers, e.g. poly(ethylene oxide) (PEO) [58], polyvinylpyrolidone (PVP) [59] and hyaluronic acid [60], or long chain DNA [61] into the Newtonian base fluid, non-Newtonian medium with elastic property can be formed. Additional normal stress difference could be introduced to exert on the suspended particles in these viscoelastic fluid. Generally, the elastic force on particles directs to the positions where the shear rate is minimum, so that three-dimensional (3D) particle focusing along channel centreline is achieved [62, 63]. Three-dimensional focusing is of great importance in on-chip flow cytometry, where particles or cells need to be aligned and spaced precisely along a single line for one-by-one detection and analysis. More recently, by combining the elastic force and inertial lift forces in a viscoelastic medium, an inertial-elastic focusing concept was introduced [58], and three-dimensional focusing of particles by employing viscoelasticity was even achieved for Reynolds number up to $Re \approx 10,000$ [60]. The mechanism of this amazing phenomenon is still unclear, which deserves further investigation. In the following sections, without specific notations, the medium is assumed to be Newtonian fluid, thus elastic force is ignored in the analysis.

3   Different Structures of Micro-channels in Inertial Microfluidics

The straight channel is often employed as a basic model to investigate the fundamental mechanism of inertial migration due to its simplicity [7, 12, 45, 64–66]. However, for applications in particle separation or sorting, the position differences for particles with different properties (size, shape, deformability) are normally not distinct enough for sufficient separation. Introduction of secondary flow by curvature or disturbance structure may amplify this difference, promoting the separation of particles by differential lateral positions. Meanwhile, the characteristics

of applied secondary flow are varying for differently structured microchannels, as well as inertial lift forces. Therefore, the number and location of final equilibrium positions of particles are highly dependent on the structure of microchannel. The characteristics of inertial forces in channels with different structures are summarized in Table 25.1.

3.1 Straight Channels

In a straight channel with circular cross section, inertial equilibrium positions are located within a narrow annulus at about 0.6 times of the channel radii form the channel centreline [1, 67]. However, in a straight channel with a rectangular cross section, the situation becomes more complex. In a square straight channel (AR = 1), particles normally focus into four symmetric equilibrium positions [68], facing the centre of each wall. Interestingly, a further reduction to two equilibrium positions happens in a low aspect ratio (AR ≈ 0.5) channel, located near the centre of long faces of the channel [12, 69, 70].

In a straight channel, particles' lateral migration velocity (U_L) and the minimum channel length (L_{min}), which is the minimum length that is required for particles to migrate to their inertial equilibrium positions, can be evaluated by balancing the net inertial lift force with the Stokes drag [52]:
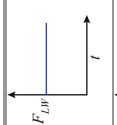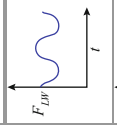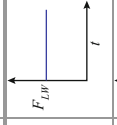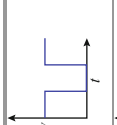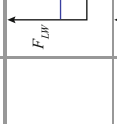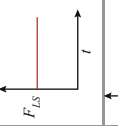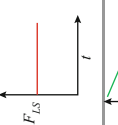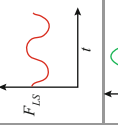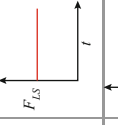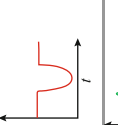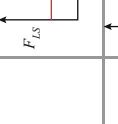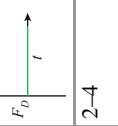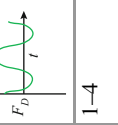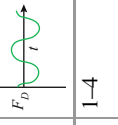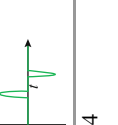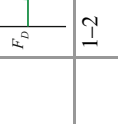
$$U_L = \frac{F_L}{3\pi\mu a} = \frac{\rho_f \overline{U}_f^2 a^3}{6\pi\mu H^2} \qquad (25.10)$$

$$L_{min} \approx \frac{H}{2U_L} * \overline{U}_f = \frac{3\pi\mu_f H^3}{\rho_f \overline{U}_f a^3} \qquad (25.11)$$

Two dimensionless Reynolds numbers can characterize the lateral migration of particles in straight channels: (1) channel Reynolds number Re describing the ratio between inertial force and viscous force of fluid and (2) particle Reynolds number $R_P \left(= Re\frac{a^2}{H^2}\right)$ additionally considers the size ratio of particle to channel. When $R_P \gg 1$, inertial migration of particles across the fluid streamlines becomes obvious [19].

Generally, straight channels are not good at particle separation. Most of the applications on particle separation by straight channels are actually based on the filtration concept, that is, larger particles are focused after a certain length due to the sufficient inertial lift forces, while their smaller counterparts remain uniformly distributed, then larger target particles can be removed from the mixture and concentrated [8, 13, 71, 72]. In order to achieve a complete separation, cascaded straight channels were applied to separate particles based on the size-dependent lateral migration speed [11].

Table 25.1 The magnitude variation of forces in microchannels with different structures [44, 113]

	Straight	Spiral	Asymmetric serpentine	Symmetric serpentine	CE array	Pillar array
Wall lift force	F_{LW}	F_{LW}	F_{LW}	F_{LW}	F_{LW}	F_{LW}
Shear gradient lift force	F_{LS}	F_{LS}	F_{LS}	F_{LS}	F_{LS}	F_{LS}
Secondary flow drag	F_D	F_D	F_D	F_D	F_D	F_D
Equilibrium positions	2–4	1–4	1–4	1–4	1–4	1–2
Theoretical separation scaling	$> a$	a^2	a^2	a^2	$\geq a^3$	NA
References	[11–14, 45, 66, 68, 71]	[22, 23, 73–76, 114]	[19, 90, 91, 93, 115]	[19, 24, 41, 88]	[31, 69, 70, 80, 100, 116–119]	[32, 85–87, 106]

Note: CE indicates contraction–expansion

3.2 Spiral Channels

Particles flowing in a curved channel with finite inertia experience both inertial lift forces and secondary flow drag. When the channel curvature is along a single constant direction, the curved channel forms a spiral channel. And the direction of secondary flow within each cross section is almost constant, even though there may be a little variation on its magnitude due to the change of channel curvature. By a first-order approximation, one can make an assumption that the effects of inertial lift force and secondary flow drag act in superposition on a particle in a spiral channel. A particle held stationary at an inertial equilibrium position experiences a secondary flow drag whose magnitude is directly proportional to the local secondary flow velocity [20]. This secondary flow drag force acts to entrain particles within the streamline of symmetrically rotating vortices, which is primarily an effect of mixing. In contrast, inertial lift forces tend to keep particles at specific equilibrium positions within channel cross section. The order of magnitude scaling between inertial lift forces and Dean drag $R_f = a^3 R/H^3$ determines the final behaviour of the suspended particles. This dimensionless parameter is useful for the prediction of particle focusing in curved channels.

At limiting conditions where (i) $R_f \rightarrow 0$, Dean drag force dominates the behaviour of particles, particles neglect inertial equilibrium positions and remain entrained within the secondary flow streamlines; and (ii) $R_f \rightarrow \infty$, inertial lift force is dominant, particles migrate to the inertial equilibrium positions independent of the secondary flow [20]. For most cases, in the intermediate range of R_f, inertial equilibrium positions can be modified by the secondary flow, leading to new intriguing focusing phenomenon. It has been noted that R_f is dependent on particle size, so that different-sized particles may occupy distinct equilibrium positions in a spiral channel. These size-dependent behaviours have been widely applied on particle separation in spiral channels [22, 23, 53, 73–79].

3.3 Straight Channels with Disturbance Structures

Besides curvature, the introduction of disturbance structures (contraction–expansion cavities [80], micro-pillar obstacles [32]) into straight channels will also induce convective secondary flow within the cross section. In the contraction–expansion array (CEA) channel, the shape of this convective secondary flow is very similar to the Dean flow in the curved/spiral channels; therefore, it was termed as Dean-like flow [31]. In contrast to the Dean flow whose magnitude is proportional to $Re^{1.63 \text{ or } 2}$ [20, 54], the magnitude of Dean-like flow was found to be linear with Re [31]. Meanwhile, the secondary flow in CEA channel only appears within the narrow contraction regions [69], and its effect on particle inertial focusing is intermittent. Furthermore, due to a suddenly increased channel width during the expansion area, the shear gradient lift force drops sharply, and the wall lift force even vanishes during this area because particles are too far from the two-sided walls.

In addition to the secondary counter-rotating flow within the cross section of the main channel around the contraction–expansion region, a horizontal micro-vortex within the contraction–expansion chamber can also be generated due to the detachment of boundary layer under a high flow speed [81]. This horizontal micro-vortex has been intensively investigated for particle selective trapping, especially for circulating tumour cells (CTCs) isolation [26, 82–84].

Besides expansion–contraction cavity, micro-pillar obstacles embedded within a micro-channel can also induce irreversible twisted flows at a finite inertial flow. Furthermore, the lateral position of the pillar can be used to tune the position and shape of the net recirculating flows. It has shown precise control of fluid transformation by programming the positions of sequenced cylindrical pillars [32, 85]. Similar to the secondary flow induced by the contraction–expansion chamber or channel curvature, the locally induced secondary flows by sequenced pillars could also be used to modify particle inertial focusing process [86, 87].

3.4   Serpentine Channels

As we know, in spiral channels, the curvature is along a single direction, therefore the secondary flow can reach steady state after a given channel length, and almost consistent within different cross sections. In a serpentine channel with alternating curves, the situation becomes more sophisticated. For example, with alternating curves, secondary flow will vary along the channel length, and particles' movement within the alternating curves may not approach steady state after each turn, and accumulation of this unsteady may cause unpredictable and non-intuitive behaviours of particles. Therefore, analysis of particle behaviour with a simple superposition method cannot always be effective.

Di Carlo et al. [19] investigated the particle inertial focusing in serpentine channels. In their study, original four equilibrium positions in a straight channel with square cross section were reduced to two in a symmetric serpentine channel. Above a critical Dean number, particle focusing was perturbed. Furthermore, in an asymmetric serpentine channel, the number of equilibrium positions could be further reduced to one. From their understanding, the balance between the inertial lift forces (F_L) and the Dean drag force (F_D) determines the preferred location of focusing positions. Dean flow does not create particle focusing, but it acts in superposition with inertial lift forces to reduce the number of inertial equilibrium positions. If $F_D \gg F_L$, no focusing will be observed; if $F_D \ll F_L$, focusing due to inertial lift forces alone will be observed. In symmetric serpentine channels, the effects of curvature-induced secondary flow will be counteracted by its corresponding opposite segments, thus secondary-flow-modified single focusing may only be achieved in the asymmetric serpentine channels [19]. However, our recent experimental and analytical studies found some discrepancies [24, 41, 88], which demand a further exploration.

4 Applications of Inertial Microfluidics

Due to the superior advantages of simplicity, high-throughput, precise manipulation and low cost, inertial microfluidics has broad applications in biomedicine (e.g. bio-fluid sample preparation and disease diagnosis) and industry (e.g. waste water purification and micro-fibre fabrication), and the application area is still expanding rapidly. Generally, its applications include but not limited to: (1) sheathless alignment; (2) particle filtration and separation; (3) solution exchange; (4) microfibre fabrication and (5) hydrodynamic phenotyping of single cell.

4.1 Sheathless Alignment

Particle alignment is essential for many biological detection systems, especially on-chip flow cytometry [89], where particles/cells are precisely positioned near an optical or electrical detector and aligned along a specific path one-by-one to achieve accurate results. In contrast to the conventional techniques utilizing sheath flow surround the sample flow to achieve 3D particle focusing in flow cytometry, sheathless inertial focusing is simple and easy to handle, without reduction in sample flow rate [44]. It has been widely explored and demonstrated in on-chip flow cytometry [55, 90–92], as shown in Fig. 25.7a. On-chip flow cytometry can also be integrated with a cell-sorting module to actively separate and purify sample subpopulations [93–95], as shown in Fig. 25.7b.

4.2 Particle Filtration and Separation

Particle filtration and separation are widely used in industry and biomedicine, such as wastewater purification, blood sample preparation and disease diagnosis. In microfluidics, continuous flow separation and sorting of particles are generally based on two principles: (1) equilibrium separation, where particles occupy different property-dependent equilibrium positions, see (I) in Fig. 25.8a; and (2) kinetic separation, where various particles possess different transport speeds perpendicular to primary flow direction under an applied force field [96, 97], as shown (II) in Fig. 25.8a. In these two separation principles, different particles occupy distinct final positions at the outlet, and complete separation can then be realized. In addition, there is another kind of mechanism depending on the forces parallel to primary flow, but it is not common in inertial microfluidics.

Particle filtration is based on the principle that target particles are focused, and the rest particles are randomly distributed, so that target particles can be get rid of particle mixtures, as shown (I) in Fig. 25.8b. However, the collected target particles are still not perfectly pure, containing partial unwanted ones. Theoretically, by
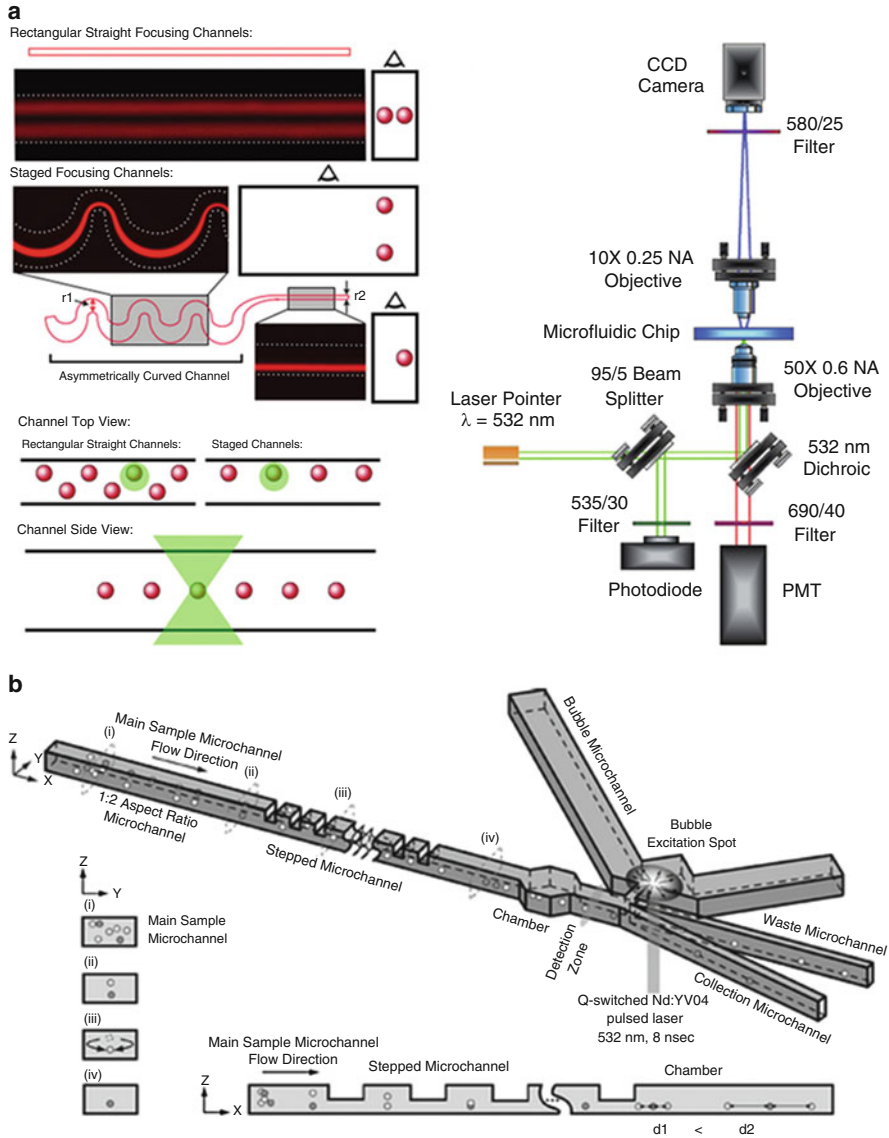
Fig. 25.7 (**a**) Schematic of on-chip flow cytometry setup. Inertial focusing in straight and asymmetric serpentine microchannels was employed to align particle along a single path one-by-one for single cell level detection downstream. Reproduced from [91]. (**b**) Inertial particle focusing in a low-aspect-ratio straight channel with a series of constrictions in height is integrated with a pulsed laser-activated cell sorter (PLACS) to realize active real-time cell sorting. Reproduced from [95]
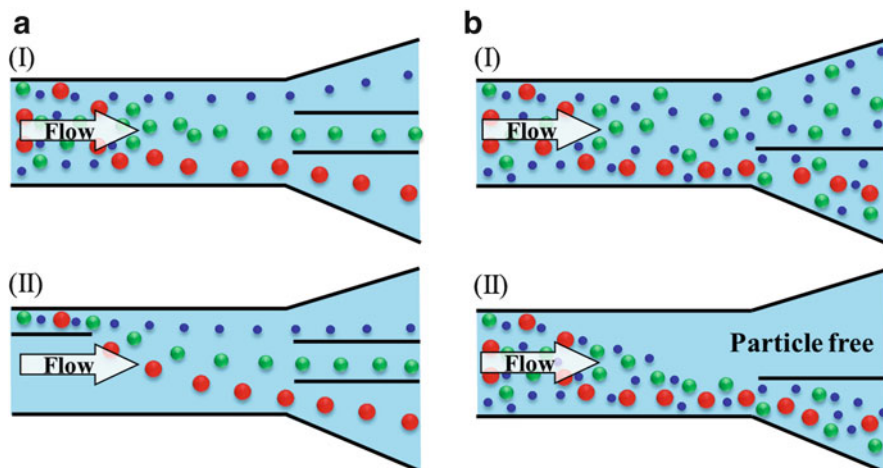
Fig. 25.8 (**a**) Two different particle separation principles in inertial microfluidics by (*I*) equilibrium separation and (*II*) kinetic separation. (**b**) Particle filtration in inertial microfluidics, (*I*) target particles are focused, and other particles remain random distributed; (*II*) all particles are focused and concentrated along one streak and removed from one branch, leaving particle-free medium

cascaded processes, target particles can be purified continuously. Another particle filtration principle is that all particles are focused along one specific path, and depleted from the medium, as shown (II) in Fig. 25.8b. It is mostly applied in the situations where the liquid phase is the target in the filtration process, such as waste water purification [98] and blood plasma separation [99].

Due to the superior advantages of simplicity, label-free and high-throughput, inertial microfluidics has great potential in routine bio-fluid preparation and disease diagnosis, especially for target cells in low abundance, e.g. circulating tumour cells (CTCs). Cancer, also known as a malignant tumour, is a group of diseases involving abnormal cell growth with the potential to invade or proliferate in other organs of the body. Metastases from primary tumours are the leading causes (∼90 %) of death for nonhematological cancers [78]. During the progression of metastasis, cancer cells escaped from solid tumours and enter the bloodstream, becoming CTCs, which hold a great potential to serve as important biomarkers for early diagnosis of cancer metastases, cancer prognosis and therapy monitoring. CTCs analyses are considered as a real-time "liquid biopsy", which is much less invasive than the current method for cancer diagnosis requiring invasive biopsy followed by molecular analyses. However, CTCs are extremely rare, comprising only a few out of one billion haematological cells in blood, making their isolation and characterization an extreme technological challenge. Due to the high-throughput feature of inertial microfluidics, it can process a larger volume of blood samples in a much shorter time than other techniques. A large number of inertial microfluidic systems for CTCs separation and isolation has been proposed and reported [75, 84, 100–102] and have shown very promising capacity in CTCs isolation.
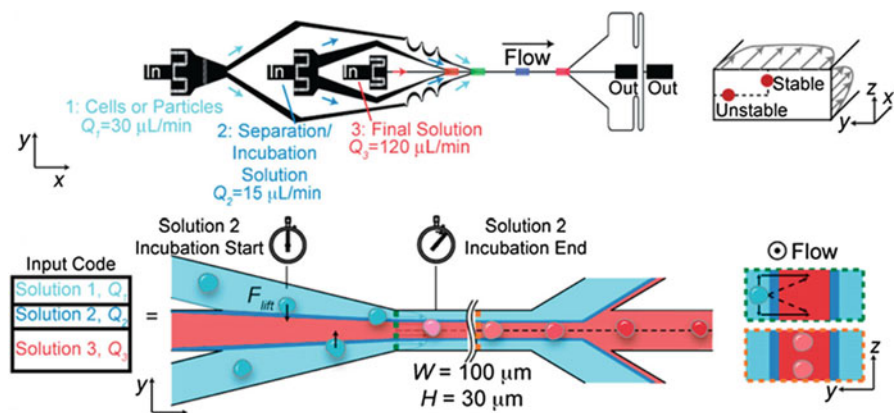
Fig. 25.9 Rapid inertial solution exchange in a straight channel for controlled interaction of reagents with cells and particles. Reproduced from [103]

4.3 Solution Exchange

Manual concentration, staining and washing procedures for the cellular sample are routinely conducted by cytopathologists as a means of diagnosing malignancies and other diseases. However, current approaches for chemical treatments of cells and chemical reactions typically operate on slow time scales (∼seconds to minutes), limiting the realm of fast molecular events. Inertial lateral migration has been subtly employed as a helpful tool to mediate millisecond reaction time around particles and cells [103], as shown in Fig. 25.9. The main transfer channel is designed with low AR, so that particles will migrate to the lateral centre of the channel regardless of their initial positions in the inlet. The channel is designed with three inlets. Particles suspension is infused from the two-sided inlets, while the reagent is delivered from the central inlet. Lateral migration of particles from original medium to the middle stream area initiates the contact of particles/cells with the chemical reagent, and duration of particles within middle stream mediates the reaction time. The proposed microfluidic system can perform several functions in the sample preparation for cytopathology that (i) automates colorimetric staining on-chip; (ii) images cells in flow and (iii) provides additional quantitative analyses of captured images [104].

4.4 Microfibre Fabrication

In the conventional fibre fabrication technology, circular cross section is the most prevalent. Although non-circular cross-sectional fibres can be produced by the standard spinning process through the non-circular hole of a spinneret, it is still challenging to form complex cross-sectional shapes in some materials because of die
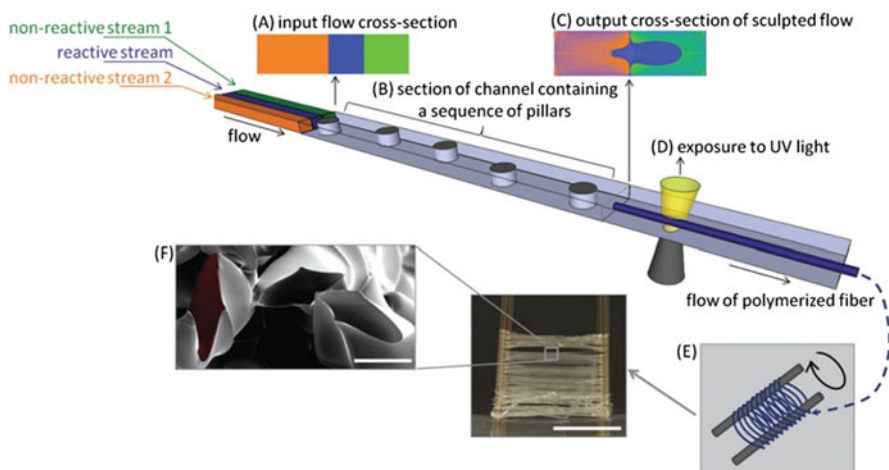
Fig. 25.10 Schematic illustration of the synthesis of shaped polymeric microfibres through sequenced micro-pillars in a programmable inertial microfluidic system. Reproduced from [85]

swelling, the rate of solidification and polymer relaxation [85, 105]. Programmable flow deformation by sequenced micro-pillars in inertial microfluidics provides a possible solution.

A micro-pillar within a micro-channel induces irreversibly twisted flows at a finite inertial flow. Furthermore, the lateral position of the pillar can be used to tune the position and shape of the net recirculating flows within the cross section. Therefore, it may enable precise control of the net fluid transformation by programming the positions of sequenced cylindrical pillars. Through a computer-aided design (CAD) tool uFlow which has a stored library of pre-computed fluid deformation of the individual pillar in the flow channel, one can design and predict the net flow deformation, without requirement of significant design iteration and experimental verification [32, 106]. Nunes et al. [85] utilized this microfluidic technique to fabricate polymeric microfibres with noncircular cross-sectional shape, as shown in Fig. 25.10. The cross-sectional shapes of various fabricated microfibres agreed reasonably well with that predicted using the uFlow code.

4.5 Hydrodynamic Stretching

The mechanical properties of cells have become promising biomarkers of cell state and disease, such as metastatic potential [107], cell cycle stage [108], degree of differentiation [109] and leukocyte activation [110]. Through a unique combination of inertial focusing, hydrodynamic stretching and automated image analysis, an automated microfluidic technology was developed to screen single cell deforma-bility at approximately 2000 cells/s [111]. Sheathless inertial focusing by the
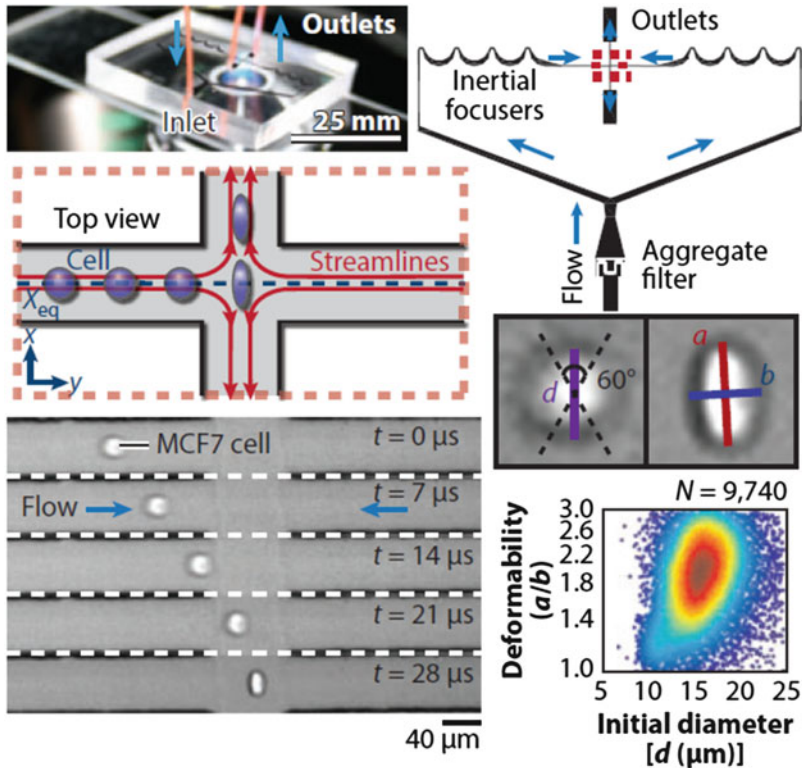
Fig. 25.11 Deformability cytometry to characterize mechanical morphology of single cells in a large population. Inertial focusing of cells in a serpentine channel orders cells along a single stream for downstream hydrodynamic pinching. Reproduced from [111]

asymmetric serpentine channel delivers suspended cells uniformly to an extensional flow region, and two sheath flows pinch the cells in the downstream, as shown in Fig. 25.11. Uniform delivery leads to uniformity in the hydrodynamic stress and corresponding deformation of individual cells. This microfluidic technology has been applied in the identification of inflammation and malignancy in pleural fluids and stem cell state [111]. By modification of extensional flow by self-feeding manner, the deformability cytometry can enable a much higher throughput (65,000 cells/s) [112].

4.6   Summary

Inertial microfluidic technology has demonstrated broad applications in biomedicine and industry, and the functionality is largely dependent on the structure of micro-channel. Table 25.2 summarizes the applications of differently structured

Table 25.2 Applications of differently structured microchannels in inertial microfluidics

Channel types	Applications	Sample	References
Straight	Flow cytometry	Particles, Diluted whole blood, 1 % or 5 % v/v RBCs	[92]
	Filtration or separation	Digestions of murine adrenal glands, 0.5 % (v/v) blood, 0.5 % (w/w) particles, *E. coli* spiked in blood	[71, 72]
		Particle mixture, CTCs spiked in diluted blood (0.8 % v/v)	[11]
		CTCs spiked in blood at 1:100, total concentration (2×10^6 cells/mL)	[16]
	Solution exchange	MCF7 and MCF10A breast cancer cell lines, HeLa cell line	[103, 104]
Spiral	Flow cytometry	Particles (0.05 % v/v), SH-SY5Y neuroblastoma cells (0.05 % v/v)	[55]
	Filtration or separation	Particles	[22, 114, 120, 121]
		Particles, SH-SY5Y neuroblastoma, glioma cells (0.05 % v/v)	[23]
		Particles, CTCs spiked in leukocytes	[122]
		Different cell cycle human mesenchymal stem cells	[77]
		CTCs spiked in whole blood, clinical cancer patient blood sample	[73–75, 79, 102, 123]
		Leukocytes from the whole blood (0.5–2 % v/v)	[76]
	Cell encapsulation	HL60, K562 (1–2.2 % v/v)	[56]

(continued)

Table 25.2 (continued)

Channel types	Applications	Sample	References
Serpentine	Flow cytometry	10.2 μm particles	[91]
		Budding yeast and MCF7 cells in blood	[90]
	Filtration or separation	Particles, diluted blood (2 % v/v)	[115]
		CTCs spiked in blood, clinical samples from metastatic cancer patients	[93]
		Diluted whole blood, MEL cells spiked in whole blood	[41, 88]
	Hydrodynamic phenotyping	Mouse and human embryonic stem cells (ESCs)	[111]
		Jurkat leukaemia and HeLa cells	[112]
Contraction–Expansion array (CEA)	Flow cytometry with pulsed laser-activated cell sorter	Particles, Ramos human Burkitt lymphoma Cells	[95]
	Filtration or separation	Parasites spiked in RBCs (2.5 % v/v)	[124]
		Particles	[69, 118, 119, 125, 126]
		Undiluted whole blood	[116]
		MCF-7 cancer cells spiked in the whole blood (∼45 % v/v) at 1:1000	[100]
		CTCs spiked in the whole blood and pleural effusions, clinical blood sample from cancer patient	[26, 83, 84]
	Solution exchange	MCF7 cells	[26]
Pillar array	Microfibre fabrication	PEG-DA	[85]
	Solution exchange	Particles, WBCS after blood lysis	[86]

Notes: particles here specifically indicate polystyrene beads

microchannels in inertial microfluidics. Due to the superior advantages of high-throughput, simplicity, precise manipulation and low cost, inertial microfluidics is very promising in bio-fluid preparation and particle separation, especially for the rare target cells isolation (e.g. circulating tumour cells or CTCs). Since target cells are low in abundance, particle suspension with large volume needs to be processed within reasonably short time. Under such situation, the advantages of inertial microfluidics are obvious.

5   Conclusions

In this chapter, we discussed three important aspects of inertial microfluidics: fundamental mechanism, microchannel designs and applications. We hope that readers can have a brief idea what the inertial microfluidics is, why the inertial focusing happens, and how the inertial focusing can be applied in biomedicine and industry. It should be noted that although extensive investigations on the mechanism of particle inertial focusing have been conducted, and a variety of applications on particle/cell separation have been demonstrated, quantitative design rules are still lacking for particle inertial manipulation in differently shaped channels. Therefore, more dedicated work is needed to uncover the detailed underlying mechanism and provide explicit rules to the end users even without specialized knowledge and experience about inertial microfluidics. In addition, more efforts are required to improve the separation resolution and processing speed through optimization of channel structure and scale up scheme.

Inertial microfluidics, as an unconventional microfluidic technology, provides more possibilities and opportunities on dealing with macroscale biofluid samples within well-defined microscale channels, and would be a promising replacement of conventional macro-scale platform (flow cytometry, centrifuge and membrane filter) for bioparticle precise processing and separation. We believe that in the near future inertial microfluidics remains a hot research topic with expanding applications, due to the incredible advantages of high-throughput, simple structure and low cost.

References

1. Segre G (1961) Radial particle displacements in Poiseuille flow of suspensions. Nature 189:209–210
2. Segre G, Silberberg A (1962) Behaviour of macroscopic rigid spheres in Poiseuille flow. Part 2. Experimental results and interpretation. J Fluid Mech 14:136–157
3. McLaughlin JB (1993) The lift on a small sphere in wall-bounded linear shear flows. J Fluid Mech 246:249–265
4. Joseph DD, Ocando D (2002) Slip velocity and lift. J Fluid Mech 454:263–286
5. Cherukat P, McLaughlin JB (1994) The inertial lift on a rigid sphere in a linear shear flow field near a flat wall. J Fluid Mech 263:1–18

6. Asmolov ES (1999) The inertial lift on a spherical particle in a plane Poiseuille flow at large channel Reynolds number. J Fluid Mech 381:63–87
7. Chun B, Ladd A (2006) Inertial migration of neutrally buoyant particles in a square duct: an investigation of multiple equilibrium positions. Phys Fluids 18:031704
8. Tanaka T, Ishikawa T, Numayama-Tsuruta K, Imai Y, Ueno H, Yoshimoto T et al (2012) Inertial migration of cancer cells in blood flow in microchannels. Biomed Microdevices 14:25–33
9. Yang BH, Wang J, Joseph DD, Hu HH, Pan T-W, Glowinski R (2005) Migration of a sphere in tube flow. J Fluid Mech 540:109–131
10. Hood K, Lee S, Roper M (2015) Inertial migration of a rigid sphere in three-dimensional Poiseuille flow. J Fluid Mech 765:452–479
11. Zhou J, Giridhar PV, Kasper S, Papautsky I (2013) Modulation of aspect ratio for complete separation in an inertial microfluidic channel. Lab Chip 13:1919–1929
12. Zhou J, Papautsky I (2013) Fundamentals of inertial focusing in microchannels. Lab Chip 13:1121–1132
13. Bhagat AAS, Kuntaegowdanahalli SS, Papautsky I (2008) Enhanced particle filtration in straight microchannels using shear-modulated inertial migration. Phys Fluids 20:101702
14. Ciftlik AT, Ettori M, Gijs MA (2013) High throughput-per-footprint inertial focusing. Small 9:2764–2773
15. Hur SC, Choi S-E, Kwon S, Di Carlo D (2011) Inertial focusing of non-spherical microparticles. Appl Phys Lett 99:044101
16. Hur SC, Henderson-Maclennan NK, McCabe ER, Di Carlo D (2011) Deformability-based cell classification and enrichment using inertial microfluidics. Lab Chip 11:912–920
17. Ho BP, Leal LG (1974) Inertial migration of rigid spheres in two-dimensional unidirectional flows. J Fluid Mech 65:365–400
18. Amini H, Lee W, Di Carlo D (2014) Inertial microfluidic physics. Lab Chip 14:2739–2761
19. Di Carlo D, Irimia D, Tompkins RG, Toner M (2007) Continuous inertial focusing, ordering, and separation of particles in microchannels. Proc Natl Acad Sci U S A 104:18892–18897
20. Di Carlo D (2009) Inertial microfluidics. Lab Chip 9:3038–3046
21. Wu Z, Willing B, Bjerketorp J, Jansson JK, Hjort K (2009) Soft inertial microfluidics for high throughput separation of bacteria from human blood cells. Lab Chip 9:1193–1199
22. Bhagat AAS, Kuntaegowdanahalli SS, Papautsky I (2008) Continuous particle separation in spiral microchannels using dean flows and differential migration. Lab Chip 8:1906–1914
23. Kuntaegowdanahalli SS, Bhagat AAS, Kumar G, Papautsky I (2009) Inertial microfluidics for continuous particle separation in spiral microchannels. Lab Chip 9:2973–2980
24. Zhang J, Li W, Li M, Alici G, Nguyen N-T (2013) Particle inertial focusing and its mechanism in a serpentine microchannel. Microfluid Nanofluid 17:305–316
25. Lim DSW, Shelby JP, Kuo JS, Chiu DT (2003) Dynamic formation of ring-shaped patterns of colloidal particles in microfluidic systems. Appl Phys Lett 83:1145–1147
26. Mach AJ, Kim JH, Arshi A, Hur SC, Di Carlo D (2011) Automated cellular sample preparation using a Centrifuge-on-a-Chip. Lab Chip 11:2827–2834
27. Zhang J, Yan S, Alici G, Nguyen N-T, Di Carlo D, Li W (2014) Real-time control of inertial focusing in microfluidics using dielectrophoresis (DEP). RSC Adv 4:62076–62085
28. Russom A, Gupta AK, Nagrath S, Di Carlo D, Edd JF, Toner M (2009) Differential inertial focusing of particles in curved low-aspect-ratio microchannels. New J Phys 11:075025
29. Berger S, Talbot L, Yao L (1983) Flow in curved pipes. Annu Rev Fluid Mech 15:461–512
30. Lee MG, Choi S, Park JK (2009) Rapid laminating mixer using a contraction-expansion array microchannel. Appl Phys Lett 95:051902
31. Zhang J, Li M, Li W, Alici G (2013) Inertial focusing in a straight channel with asymmetrical expansion–contraction cavity arrays using two secondary flows. J Micromech Microeng 23:085023
32. Amini H, Sollier E, Masaeli M, Xie Y, Ganapathysubramanian B, Stone HA et al (2013) Engineering fluid flow using sequenced microstructures. Nat Commun 4:1826

33. Çetin B, Li D (2011) Dielectrophoresis in microfluidics technology. Electrophoresis 32: 2410–2427
34. Li M, Li S, Li W, Wen W, Alici G (2013) Continuous manipulation and separation of particles using combined obstacle- and curvature-induced direct current dielectrophoresis. Electrophoresis 34:952–960
35. Forbes TP, Forry SP (2012) Microfluidic magnetophoretic separations of immunomagneti-cally labeled rare mammalian cells. Lab Chip 12:1471–1479
36. Shen F, Hwang H, Hahn YK, Park JK (2012) Label-free cell separation using a tunable magnetophoretic repulsion force. Anal Chem 84:3075–3081
37. Li S, Ding X, Guo F, Chen Y, Lapsley MI, Lin S-CS et al (2013) An on-chip, multichannel droplet sorter using standing surface acoustic waves (SSAW). Anal Chem 85:5468–5474
38. Destgeer G, Lee KH, Jung JH, Alazzam A, Sung HJ (2013) Continuous separation of particles in a PDMS microfluidic channel via travelling surface acoustic waves (TSAW). Lab Chip 13:4210–4216
39. MacDonald M, Spalding G, Dholakia K (2003) Microfluidic sorting in an optical lattice. Nature 426:421–424
40. Jung JH, Lee KH, Lee KS, Ha BH, Oh YS, Sung HJ (2014) Optical separation of droplets on a microfluidic platform. Microfluid Nanofluid 16:635–644
41. Zhang J, Yan S, Sluyter R, Li W, Alici G, Nguyen N-T (2014) Inertial particle separation by differential equilibrium positions in a symmetrical serpentine micro-channel. Sci Rep 4: Art No. 4527
42. Michaelides E (2006) Particles, bubbles & drops: their motion, heat and mass transfer. World Scientific, Singapore
43. Saffman PG (1965) The lift on a small sphere in a slow shear flow. J Fluid Mech 22:385–400
44. Martel JM, Toner M (2014) Inertial focusing in microfluidics. Annu Rev Biomed Eng 16:371–396
45. Di Carlo D, Edd JF, Humphry KJ, Stone HA, Toner M (2009) Particle segregation and dynamics in confined flows. Phys Rev Lett 102:94503
46. Richardson JF, Coulson JM, Harker J, Backhurst J (2002) Chemical engineering: particle technology and separation processes, vol 2. Butterworth-Heinemann, Oxford
47. Rubinow S, Keller JB (1961) The transverse force on a spinning sphere moving in a viscous fluid. J Fluid Mech 11:447–459
48. Matas J, Morris J, Guazzelli E (2004) Lateral forces on a sphere. Oil Gas Sci Technol 59:59–70
49. Kim YW, Yoo JY (2009) Axisymmetric flow focusing of particles in a single microchannel. Lab Chip 9:1043–1045
50. Vasseur P, Cox RG (1977) The lateral migration of spherical particles sedimenting in a stagnant bounded fluid. J Fluid Mech 80:561–591
51. Feng J, Hu HH, Joseph DD (1994) Direct simulation of initial value problems for the motion of solid bodies in a Newtonian fluid. Part 2. Couette and Poiseuille flows. J Fluid Mech 277:271–301
52. Bhagat AAS, Kuntaegowdanahalli SS, Papautsky I (2009) Inertial microfluidics for continu-ous particle filtration and extraction. Microfluid Nanofluid 7:217–226
53. Guan G, Wu L, Bhagat AA, Li Z, Chen PC, Chao S et al (2013) Spiral microchannel with rectangular and trapezoidal cross-sections for size based particle separation. Sci Rep 3: Art No. 1475
54. Squires TM, Quake SR (2005) Microfluidics: fluid physics at the nanoliter scale. Rev Mod Phys 77:977
55. Bhagat AAS, Kuntaegowdanahalli SS, Kaval N, Seliskar CJ, Papautsky I (2010) Inertial microfluidics for sheath-less high-throughput flow cytometry. Biomed Microdevices 12:187–195
56. Kemna EW, Schoeman RM, Wolbers F, Vermes I, Weitz DA, van den Berg A (2012) High-yield cell ordering and deterministic cell-in-droplet encapsulation using Dean flow in a curved microchannel. Lab Chip 12:2881–2887

57. Masaeli M, Sollier E, Amini H, Mao W, Camacho K, Doshi N et al (2012) Continuous inertial focusing and separation of particles by shape. Phys Rev X 2:031017

58. Yang S, Kim JY, Lee SJ, Lee SS, Kim JM (2011) Sheathless elasto-inertial particle focusing and continuous separation in a straight rectangular microchannel. Lab Chip 11:266–273

59. Seo KW, Kang YJ, Lee SJ (2014) Lateral migration and focusing of microspheres in a microchannel flow of viscoelastic fluids. Phys Fluids 26:063301

60. Lim EJ, Ober TJ, Edd JF, Desai SP, Neal D, Bong KW et al (2014) Inertio-elastic focusing of bioparticles in microchannels at high throughput. Nat Commun 5: Art No. 4120

61. Kang K, Lee SS, Hyun K, Lee SJ, Kim JM (2013) DNA-based highly tunable particle focuser. Nat Commun 4: Art No. 2567

62. D'Avino G, Romeo G, Villone MM, Greco F, Netti PA, Maffettone PL (2012) Single line particle focusing induced by viscoelasticity of the suspending liquid: theory, experiments and simulations to design a micropipe flow-focuser. Lab Chip 12:1638–1645

63. Del Giudice F, Romeo G, D'Avino G, Greco F, Netti PA, Maffettone PL (2013) Particle alignment in a viscoelastic liquid flowing in a square-shaped microchannel. Lab Chip 13:4263–4271

64. Amini H, Sollier E, Weaver WM, Di Carlo D (2012) Intrinsic particle-induced lateral transport in microchannels. Proc Natl Acad Sci U S A 109:11593–11598

65. Lee W, Amini H, Stone HA, Di Carlo D (2010) Dynamic self-assembly and control of microfluidic particle crystals. Proc Natl Acad Sci U S A 107:22413–22418

66. Liu C, Hu G, Jiang X, Sun J (2015) Inertial focusing of spherical particles in rectangular microchannels over a wide range of Reynolds numbers. Lab Chip 15:1168–1177

67. Matas J-P, Morris JF, Guazzelli É (2004) Inertial migration of rigid spherical particles in Poiseuille flow. J Fluid Mech 515:171–195

68. Choi Y-S, Seo K-W, Lee S-J (2011) Lateral and cross-lateral focusing of spherical particles in a square microchannel. Lab Chip 11:460–465

69. Lee MG, Choi S, Park JK (2011) Inertial separation in a contraction–expansion array microchannel. J Chromatogr A 1218:4138–4143

70. Chung AJ, Gossett DR, Di Carlo D (2013) Three dimensional, sheathless, and high-throughput microparticle inertial focusing through geometry-induced secondary flows. Small 9:685–690

71. Hur SC, Brinckerhoff TZ, Walthers CM, Dunn JC, Di Carlo D (2012) Label-free enrichment of adrenal cortical progenitor cells using inertial microfluidics. PLoS One 7:e46550

72. Mach AJ, Di Carlo D (2010) Continuous scalable blood filtration device using inertial microfluidics. Biotechnol Bioeng 107:302–311

73. Sun J, Li M, Liu C, Zhang Y, Liu D, Liu W et al (2012) Double spiral microchannel for label-free tumor cell separation and enrichment. Lab Chip 12:3952–3960

74. Sun J, Liu C, Li M, Wang J, Xianyu Y, Hu G et al (2013) Size-based hydrodynamic rare tumor cell separation in curved microfluidic channels. Biomicrofluidics 7:011802

75. Warkiani ME, Guan G, Luan KB, Lee WC, Bhagat AAS, Chaudhuri PK et al (2014) Slanted spiral microfluidics for the ultra-fast, label-free isolation of circulating tumor cells. Lab Chip 14:128–137

76. Wu L, Guan G, Hou HW, Bhagat AAS, Han J (2012) Separation of leukocytes from blood using spiral channel with trapezoid cross-section. Anal Chem 84:9324–9331

77. Lee WC, Bhagat AAS, Huang S, Van Vliet KJ, Han J, Lim CT (2011) High-throughput cell cycle synchronization using inertial forces in spiral microchannels. Lab Chip 11:1359–1367

78. Sheng W, Chen T, Kamath R, Xiong X, Tan W, Fan ZH (2012) Aptamer-enabled efficient isolation of cancer cells from whole blood using a microfluidic device. Anal Chem 84:4199–4206

79. Hou HW, Warkiani ME, Khoo BL, Li ZR, Soo RA, Tan DS-W et al (2013) Isolation and retrieval of circulating tumor cells using centrifugal forces. Sci Rep 3: Art No. 1259

80. Lee MG, Choi S, Park JK (2009) Three-dimensional hydrodynamic focusing with a single sheath flow in a single-layer microfluidic device. Lab Chip 9:3155–3160

81. Shelby JP, Lim DS, Kuo JS, Chiu DT (2003) Microfluidic systems: high radial acceleration in microvortices. Nature 425:38
82. Hur SC, Mach AJ, Di Carlo D (2011) High-throughput size-based rare cell enrichment using microscale vortices. Biomicrofluidics 5:022206
83. Che J, Mach AJ, Go DE, Talati I, Ying Y, Rao J et al (2013) Microfluidic purification and concentration of malignant pleural effusions for improved molecular and cytomorphological diagnostics. PLoS One 8:e78194
84. Sollier E, Go DE, Che J, Gossett DR, O'Byrne S, Weaver WM et al (2014) Size-selective collection of circulating tumor cells using Vortex technology. Lab Chip 14:63–77
85. Nunes JK, Wu CY, Amini H, Owsley K, Di Carlo D, Stone HA (2014) Fabricating shaped microfibers with inertial microfluidics. Adv Mater 26:3712–3717
86. Sollier E, Amini H, Go D, Sandoz P, Owsley K, Di Carlo D (2015) Inertial microfluidic programming of microparticle-laden flows for solution transfer around cells and particles. Microfluid Nanofluid 19:53–65
87. Chung AJ, Pulido D, Oka JC, Amini H, Masaeli M, Di Carlo D (2013) Microstructure-induced helical vortices allow single-stream and long-term inertial focusing. Lab Chip 13:2942–2949
88. Zhang J, Yan S, Li W, Alici G, Nguyen N-T (2014) High throughput extraction of plasma using a secondary flow-aided inertial microfluidic device. RSC Adv 4:33149–33159
89. Mao X, Lin S-CS, Dong C, Huang TJ (2009) Single-layer planar on-chip flow cytometer using microfluidic drifting based three-dimensional (3D) hydrodynamic focusing. Lab Chip 9:1583–1589
90. Goda K, Ayazi A, Gossett DR, Sadasivam J, Lonappan CK, Sollier E et al (2012) High-throughput single-microparticle imaging flow analyzer. Proc Natl Acad Sci U S A 109:11630–11635
91. Oakey J, Applegate RW Jr, Arellano E, Carlo DD, Graves SW, Toner M (2010) Particle focusing in staged inertial microfluidic devices for flow cytometry. Anal Chem 82:3862–3867
92. Hur SC, Tse HTK, Di Carlo D (2010) Sheathless inertial cell ordering for extreme throughput flow cytometry. Lab Chip 10:274–280
93. Ozkumur E, Shah AM, Ciciliano JC, Emmink BL, Miyamoto DT, Brachtel E et al (2013) Inertial focusing for tumor antigen–dependent and–independent sorting of rare circulating tumor cells. Sci Transl Med 5: 179ra47
94. Gawad S, Valero A, Braschler T, Holmes D, Renaud P (2012) On-chip flow cytometry. In: Encyclopedia of nanotechnology. Springer, pp 1913–1923
95. Chen Y, Chung AJ, Wu TH, Teitell MA, Di Carlo D, Chiou PY (2014) Pulsed laser activated cell sorting with three dimensional sheathless inertial focusing. Small 10:1746–1751
96. Gossett DR, Weaver WM, Mach AJ, Hur SC, Tse HTK, Lee W et al (2010) Label-free cell separation and sorting in microfluidic systems. Anal Bioanal Chem 397:3249–3267
97. Pamme N (2007) Continuous flow separations in microfluidic devices. Lab Chip 7:1644–1659
98. PARC. (16 Oct). Low-energy, compact, cost-effective separation for drinking water and wastewater treatment along with precious resource recovery. http://www.parc.com/work/focus-area/clean-water/
99. Kersaudy-Kerhoas M, Sollier E (2013) Micro-scale blood plasma separation: from acoustophoresis to egg-beaters. Lab Chip 13:3323–3346
100. Lee MG, Shin JH, Bae CY, Choi S, Park J-K (2013) Label-free cancer cell separation from human whole blood using inertial microfluidics at low shear stress. Anal Chem 85:6213–6218
101. Moon HS, Kwon K, Kim SI, Han H, Sohn J, Lee S et al (2011) Continuous separation of breast cancer cells from blood samples using multi-orifice flow fractionation (MOFF) and dielectrophoresis (DEP). Lab Chip 11:1118–1125
102. Khoo BL, Warkiani ME, Tan DS-W, Bhagat AAS, Irwin D, Lau DP et al (2014) Clinical validation of an ultra high-throughput spiral microfluidics for the detection and enrichment of viable circulating tumor cells. PLoS One 9:e99409

103. Dudani JS, Go DE, Gossett DR, Tan AP, Di Carlo D (2014) Mediating millisecond reaction time around particles and cells. Anal Chem 86:1502–1510

104. Tan AP, Dudani JS, Arshi A, Lee RJ, Henry T, Gossett DR et al (2014) Continuous-flow cytomorphological staining and analysis. Lab Chip 14:522–531

105. Jung I, Kim SY, Oh TH (2010) Effects of spinning conditions on shape changes of trilobal-shaped fibers. Text Res J 80:12–18

106. Stoecklein D, Wu C-Y, Owsley K, Xie Y, Di Carlo D, Ganapathysubramanian B (2014) Micropillar sequence designs for fundamental inertial flow transformations. Lab Chip 14:4197–4204

107. Suresh S, Spatz J, Mills J, Micoulet A, Dao M, Lim C et al (2005) Connections between single-cell biomechanics and human disease states: gastrointestinal cancer and malaria. Acta Biomater 1:15–30

108. Théry M, Bornens M (2008) Get round and stiff for mitosis. HFSP J 2:65–71

109. Chowdhury F, Na S, Li D, Poh Y-C, Tanaka TS, Wang F et al (2010) Material properties of the cell dictate stress-induced spreading and differentiation in embryonic stem cells. Nat Mater 9:82–88

110. Khismatullin DB (2009) The cytoskeleton and deformability of white blood cells. Curr Top Membr 64:47–111

111. Gossett DR, Henry T, Lee SA, Ying Y, Lindgren AG, Yang OO et al (2012) Hydrodynamic stretching of single cells for large population mechanical phenotyping. Proc Natl Acad Sci U S A 109:7630–7635

112. Dudani JS, Gossett DR, Henry T, Di Carlo D (2013) Pinched-flow hydrodynamic stretching of single-cells. Lab Chip 13:3728–3734

113. Zhang J, Yan S, Yuan D, Alici G, Nguyen N-T, Ebrahimi Warkiani M et al (2016) Fundamentals and applications of inertial microfluidics: a review. Lab Chip 16:10–34

114. Yoon DH, Ha JB, Bahk YK, Arakawa T, Shoji S, Go JS (2008) Size-selective separation of micro beads by utilizing secondary flow in a curved rectangular microchannel. Lab Chip 9:87–90

115. Di Carlo D, Jon F, Irimia D, Tompkins RG, Toner M (2008) Equilibrium separation and filtration of particles using differential inertial focusing. Anal Chem 80:2204–2211

116. Lee MG, Choi S, Kim HJ, Lim HK, Kim JH, Huh N et al (2011) Inertial blood plasma separation in a contraction–expansion array microchannel. Appl Phys Lett 98:253702

117. Park JS, Song SH, Jung HI (2009) Continuous focusing of microparticles using inertial lift force and vorticity via multi-orifice microfluidic channels. Lab Chip 9:939–948

118. Park J-S, Jung H-I (2009) Multiorifice flow fractionation: continuous size-based separation of microspheres using a series of contraction/expansion microchannels. Anal Chem 81:8280–8288

119. Sim TS, Kwon K, Park JC, Lee J-G, Jung H-I (2011) Multistage-multiorifice flow fractionation (MS-MOFF): continuous size-based separation of microspheres using multiple series of contraction/expansion microchannels. Lab Chip 11:93–99

120. Seo J, Lean MH, Kole A (2007) Membrane-free microfiltration by asymmetric inertial migration. Appl Phys Lett 91:033901

121. Seo J, Lean MH, Kole A (2007) Membraneless microseparation by asymmetry in curvilinear laminar flows. J Chromatogr A 1162:126–131

122. Kim TH, Yoon HJ, Stella P, Nagrath S (2014) Cascaded spiral microfluidic device for deterministic and high purity continuous separation of circulating tumor cells. Biomicrofluidics 8:064117

123. Warkiani ME, Khoo BL, Tan DS-W, Bhagat AAS, Lim W-T, Yap YS et al (2014) An ultra-high-throughput spiral microfluidic biochip for the enrichment of circulating tumor cells. Analyst 139:3245–3255

124. Warkiani ME, Tay AKP, Khoo BL, Xiaofeng X, Han J, Lim CT (2015) Malaria detection using inertial microfluidics. Lab Chip 15:1101–1109
125. Wang X, Zhou J, Papautsky I (2013) Vortex-aided inertial microfluidic device for continuous particle separation with high size-selectivity, efficiency, and purity. Biomicrofluidics 7:044119
126. Wang X, Papautsky I (2015) Size-based microfluidic multimodal microparticle sorter. Lab Chip 15:1350–1359

Chapter 26
Force Sensing for Micro/Meso Milling

Yu Hui Feng and Goldie Nejat

Abstract Miniaturized micro/meso-scaled 3-D components are in high demand for a variety of different applications ranging from medical devices to automotive systems. These components can be fabricated using micro/meso-milling techniques. The investigation of the cutting forces exerted during the milling process is crucial to understanding the cutting mechanism of the micro/meso-milling processes, improving overall efficiency of the micro/meso-milling operations, and monitoring tool wear and failure. However, due to the miniaturization of the material removal process at high cutting rates, measuring the characteristics of the cutting forces can be a challenging task. Piezoelectric force sensors with high resolutions and natural frequencies have been developed to address this challenge. This chapter provides a detailed overview of the different types of piezoelectric force sensors developed, the dynamic calibration techniques that have been used to calibrate these sensors as well as the direct implementation and application of these sensors for measuring cutting forces during micro/meso-milling operations.

1 Introduction

Miniaturized 3-D components within microscale size (100 nm to 100 μm in length) and meso-scale size (>100 μm in length) [1] are increasingly in demand for a variety of applications in the electronics, automotive, biotechnology, and medical industries [2]. For example, such components are implemented in cardiovascular and in vitro diagnostic devices, medical implants, portable electronic devices, cameras,

Y.H. Feng (✉) • G. Nejat

Autonomous Systems and Biomechatronics Laboratory, Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, ON, Canada M5S 3G8
e-mail: sean.feng@mail.utoronto.ca; nejat@mie.utoronto.ca

wireless devices, microscale fuel cells, fluidic micro-chemical reactors, microscale pumps, etc. [3]. In order to fabricate these complex micro/meso-scale components, micro/meso-milling machining techniques are used. Micro/meso-milling machining offers many advantages over traditional milling operations including high precision, smaller footprint, and lower cost and energy consumption [4]. The understanding of the cutting process used during the micro/meso-milling operations can provide vital information for the design and the control of the machining process in terms of efficiency improvement, parameter optimization, and tool wear reduction [5]. Namely, cutting forces can be measured using force sensors to provide feedback during the machining process as well as for analyzing the characteristics of the cutting forces [6].

Due to the miniaturization of the material removal process, the characteristics of the cutting forces are different from that of the conventional milling process [5]. The Merchant's sharp-edge cutting theorem [7] that is used in macro-cutting force predictions, which uses force equilibrium to approximate cutting forces, cannot be utilized for the micro/meso-milling operations due to the high negative rake angle and elastoplastic effects that are a result of the reduction in the scale of the milling process [5]. Efficient material removal in micro/meso-milling applications requires high rotational speeds of the spindle, whereas the magnitude of the cutting forces is considerably smaller than conventional macro-scale machining due to the size reduction [5]. The frequency of the cutting forces in micro/meso-milling application can be up to 8.3 kHz as the typical spindle speed can be as high as 250,000 rpm [8]. The typical magnitude of the cutting forces is of a few tens of milli-Newtons [9] and can be as high as 10–15 N [10]. As a result, these characteristics need to be taken into consideration when developing the force sensors for micro/meso-milling applications.

Piezoelectric force sensors are an appropriate choice of sensors to be used for measuring cutting forces at the aforementioned high frequencies within the required sensing range during micro/meso-milling operations, due to their high stiffness and sensitivity. These sensors can be implemented at two locations for measuring cutting forces: (1) in the spindle, as shown in Fig. 26.1a; or (2) under the workpiece, as shown in Fig. 26.1b. For direct measurements of the cutting forces, piezoelectric force sensors located under the workpiece are the most commonly used type of sensor [6].

This chapter provides a detailed overview of piezoelectric force sensors that have been developed, calibrated, and implemented for micro/meso-milling applications. Section 2 presents the force sensors that have been developed to address the challenges of machining at the micro/meso-scale, and Section 3 presents the dynamic calibration techniques used to calibrate these sensors. In Section 4, case studies of applications of force sensors that have been integrated for micro/meso-milling operations are discussed. The last section of the chapter presents concluding remarks and future work.

Fig. 26.1 Implementation of piezoelectric force sensors in micro/meso-milling operation: (**a**) in the spindle and (**b**) under the workpiece
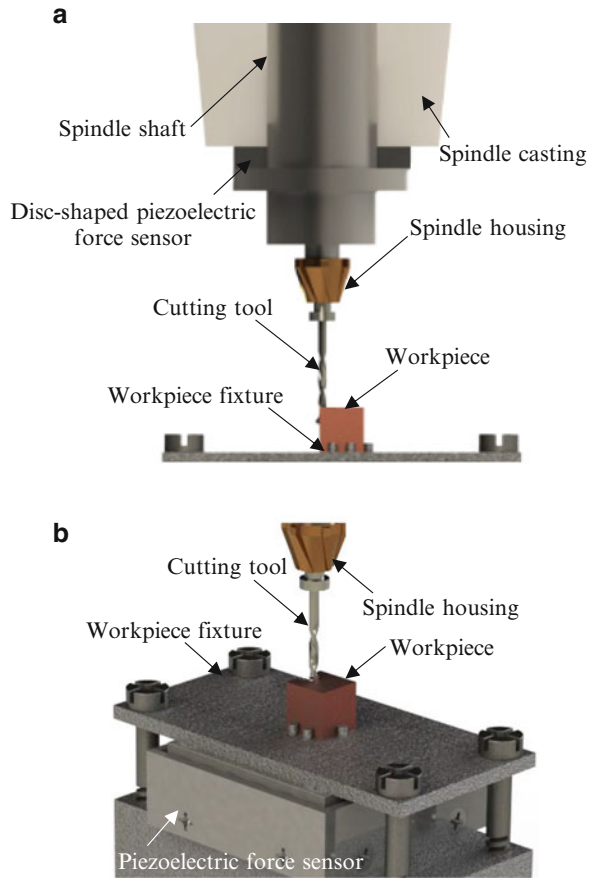
2 Piezoelectric Force Sensors for Micro/Meso Milling

A piezoelectric force sensor operates based on the piezoelectric effect, where electric charges are generated when mechanical stress is applied to the crystalline material of the sensor [11]. As the amount of the charge emitted is a function of the magnitude of the applied force at the force sensor, the force applied can be measured by measuring the charge output from the force sensor. This effect exists in various materials including natural crystals such as quartz, poled manmade ceramics such as Lead Zirconium Titanate (PZT), and some polymers such as polyvinylidenedifluoride (PVDF) [11]. These materials are used as the sensing elements of the force sensor. Their high stiffness results in high natural frequency and thus, provides the higher frequency bandwidth needed for cutting force measurements during micro/meso milling. Piezoelectric force sensors are composed of a single load cell consisting of multiple sensing elements to measure forces in the *x*, *y,* and *z* directions, Fig. 26.2a, or a dynamometer, which commonly consists of a combination of these load cells, Fig. 26.2b.

Fig. 26.2 Examples of piezoelectric force sensors: (**a**) a load cell and (**b**) a dynamometer
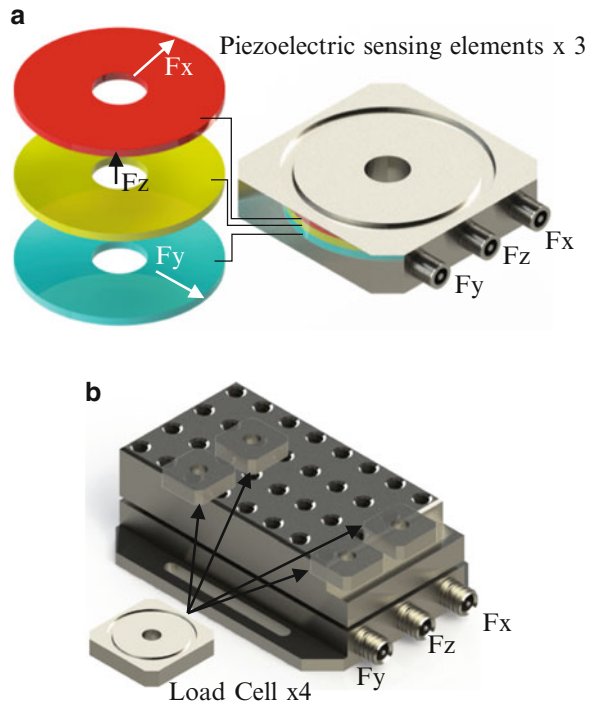
Table 26.1 Summary of force sensors for micro/meso milling

Force sensor	Resolution (mN)	Maximum measurement [N]	Maximum frequency [Hz]
Kistler 9265	10	15 k	500
Kistler 9016/9017/9018	10	1 k	1 k
Kistler 9251	10	2.5 k	3 k
Kistler high frequency dynamometer	2	250	3.67 k
MicroDyn	10	1 k	5.1 k
Micro-dynamometer	10	50	N/A
ASB Lab piezoelectric force sensor	3	10	9 k

Existing piezoelectric force sensors which meet the required specifications for measuring cutting forces during micro/meso-milling operations are summarized in Table 26.1 and are discussed in detail below. They can be categorized into two groups based on the material used for their sensing elements: (1) piezoelectric force sensors with quartz-based sensing elements and (2) piezoelectric force sensors with ceramic-based sensing elements.

2.1   Quartz-Based Piezoelectric Force Sensors

Quartz crystals have high mechanical strength and low hysteresis behavior and therefore have been a popular choice as the material used for the sensing elements of a number of commercially available force sensors [12].

The Kistler 9265 is a dynamometer consisting of four three-component load cells placed at each of the four corners of the dynamometer [13]. Within each load cell, there are three pairs of quartz sensing elements. One pair of elements is used to measure the normal force, and the other two are used to measure forces in each of the two shear directions. The specified resolution of the dynamometer is 10 mN, and the sensing range of the dynamometer is up to 15 kN in the shear directions and up to 30 kN in the normal direction [13]. The operational frequency bandwidth of the dynamometer for milling operations is 833 Hz for the normal direction and 500 Hz for both of the shear directions according to the specified natural frequencies [13].

The Kistler 901x family (9016/9017/9018) of piezoelectric force sensors consists of three pairs of ring-shaped quartz sensing elements, one for the force measurement in the normal direction and two for the force measurements in the two shear directions [14]. The resolution of the force sensor is 10 mN, whereas its measuring range is up to 2 kN in the normal direction and up to 1 kN in the shear directions [14]. In Chae and Park [15], the operational frequency bandwidth of the Kistler 9017 piezoelectric force sensor was determined using an impact hammer test. It was found that the natural frequencies in all three directions are approximately 3 kHz and the operational frequency bandwidth of the force sensor is up to 1 kHz when 10 % error tolerance is allowed.

In Kang et al. [16], an implementation of the force sensor, Kistler 9251, was presented. The force sensor contains three pairs of quartz rings mounted between two plates made of steel and protected by a welded stainless steel sensor housing. The resolution of the force sensor is 10 mN and the sensing ranging is up to 2.5 kN in the shear directions and up to 5 kN in the normal direction [17]. By conducting an impact hammer test, the natural frequency of the dynamometer was determined to be approximately 9 kHz for the normal direction and 12 kHz for both of the shear directions, indicating that the overall frequency bandwidth of the force sensor is up to 3 kHz [16].

In Garzon et al. [18], the Kistler high frequency dynamometer (HFD) was presented. The dynamometer contains a three-component quartz load cell capable of measuring the forces in the normal as well as both of the shear directions. The resolution of the force sensor is 2 mN and the measuring range is up to 250 N in all three directions [19]. To increase the natural frequency of the HFD, the structure placed between the dynamometer and the workpiece was minimized. The frequency bandwidth of the dynamometer was then determined via conducting excitation tests with an impact hammer. The results of the tests indicated that the natural frequency of the sensor is approximately 9 kHz in the normal direction and 25 kHz

in the shear directions. Therefore, the upper limit of the operational frequency bandwidth is 3.67 kHz indicating that the dynamometer can be implemented for milling operations, where a two-edge cutting tool is used with a spindle speed of 110,000 rpm.

The design and development of a piezoelectric dynamometer, MicroDyn, for high frequency force measurements in micro-machining operations was presented in Transchel et al. [20]. The MicroDyn dynamometer consists of two Kistler 9017 load cells. Each load cell was fixed onto the two opposite sides of a cuboid workpiece fixture using a base plate and preloading bolts, respectively. The dynamic behavior of the dynamometer was analyzed using an impact hammer test. The frequency response of the tests showed that the lowest natural frequency of 15.3 kHz occurred in the normal direction. Therefore, the maximum applicable excitation frequency, which is the upper limit of the operational frequency bandwidth of the dynamometer, is 5.1 kHz, corresponding to a spindle speed of 153,000 rpm for a micro-milling operation, where a two-edge cutting tool is used.

In Mekid [21], the design of a micro-dynamometer was presented. The dynamometer consists of five load cells fixed on each side of a cuboid workpiece fixture except for the top surface on which the workpiece can be affixed. The dynamometer was designed to have a sensing range of up to 50 N in both of its normal and shear directions and a resolution of 10 mN.

2.2   Ceramic Piezoelectric Force Sensor

Ceramic materials, especially PZT ceramics, are easier and cheaper to mass produce than quartz, as the high pressure environment required for quartz growth is not needed for ceramic fabrication [12]. Furthermore, ceramics have a much higher piezoelectric sensitivity (in the magnitude of 100 pC/N) compared to quartz (in the magnitude of 1 pC/N), which make them a feasible choice as the sensing material for piezoelectric force sensors [12].

In Feng et al. [22] and Feng [23], a three-component force sensor prototype, the ASB Lab piezoelectric force sensor, using piezoelectric ceramic material was developed for micro/meso-milling applications. The sensor consists of three groups of square-shaped piezoelectric ceramic sensing elements. Two sensing groups have two shear-mode sensing elements used to measure forces in each of the two shear directions, whereas the third sensing group has two compression-mode sensing elements utilized for measuring normal forces. Calibration tests were conducted in Feng et al. [24] and Feng [23] using a piezoelectric actuator to provide oscillating forces to the sensor in order to determine its specifications. The tests verified that a minimum resolution of 3 mN in all three directions, a frequency bandwidth of up to 9 kHz and a sensing range of 0–10 N was achievable for the sensor.

3 Dynamic Calibration of Force Sensors

As the cutting forces measured during the micro/meso-milling operations oscillate at high frequencies, piezoelectric force sensors need to be calibrated using dynamic calibration methods [25]. These existing dynamic calibration methods can be divided into three categories based on the characteristics of the force input [25]: (1) methods using oscillating input forces, (2) methods using impulse input forces, and (3) methods using step input forces.

3.1 Dynamic Calibration Methods Using Oscillating Forces

In Kumme [26], a calibration method using oscillating forces was introduced. By using this method, a force sensor can be calibrated with a reference sensor that has well-known dynamic properties. The calibration setup shown in Fig. 26.3 was utilized. The setup consisted of the force sensor under calibration, a rigid cylinder with known weight, and a reference force sensor. The cylinder was fixed between the two sensors and placed onto a shaker utilized to provide vibration forces. The entire system was placed within a load frame for the calibration process. According to the mass-spring-damper model of the setup, the difference between the measurements from the force sensor under calibration and the measurements from the reference sensor equals to the inertia force of the mass between the two force sensors. The weight of the mass was the combined weights of the cylinder and the effective masses of both sensors attached to the cylinder. During the calibration process, harmonic excitation was induced to the system by the shaker. Force readings were recorded by both force sensors. A single-axial accelerometer was affixed onto the cylinder to measure its acceleration. The inertia force was calculated as the
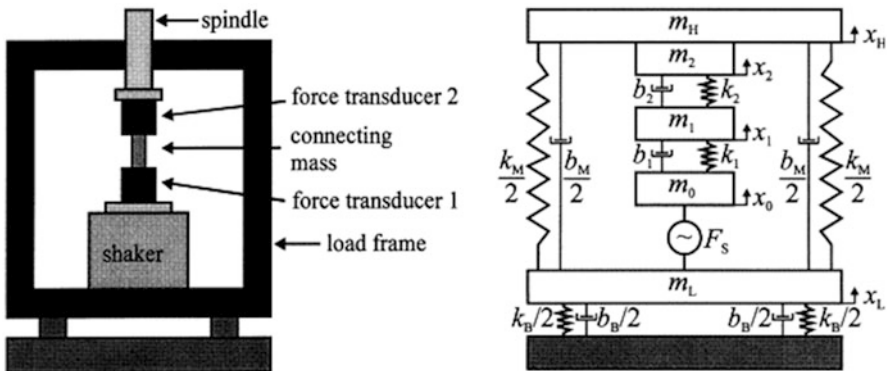
Fig. 26.3 Calibration setup and corresponding mass-spring-damper model in Kumme [26] (Source: Kumme, R. (1998), © Elsevier)

Fig. 26.4 Oscillating force calibration setup in Park et al. [27] using a shaker. a_a, a_b and a_m represent accelerometers at the top of the adaptor; base of the adaptor and on the prism, respectively (Source: Park, Y. K., Kumme, R., & Kang, D. I. (2002),© IOP Publishing)
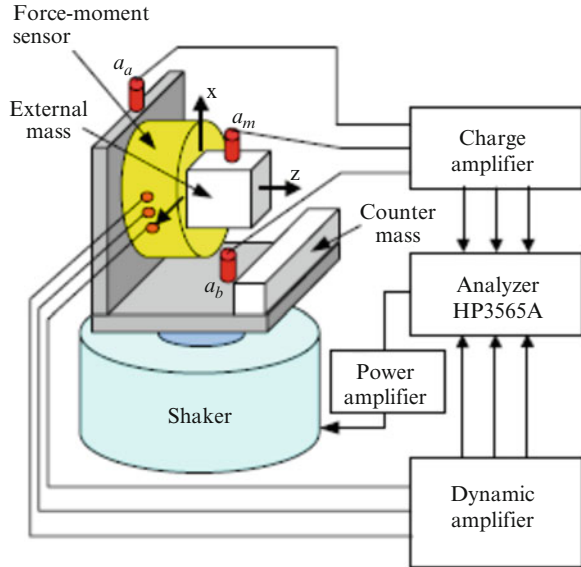
product of the measurement from the accelerometer and the weight of the mass. The frequency of the oscillation was increased from 20 to 1000 Hz with an increment of 10 Hz. The amplitude of the force was fixed at 500 N. The readings from the reference force sensor and the inertia force estimation from the accelerometer readings were then used as the baseline to compare with the measurement from the force sensor under calibration.

In Park et al. [27], the calibration of a three-component force-moment sensor using oscillating forces was presented. The calibration setup is shown in Fig. 26.4. The force-moment sensor utilized strain gauges to measure forces in both transverse directions and the moment in the normal direction. The oscillation was generated by a shaker. For calibration of each of the force-components, the force sensor was mounted onto the shaker with an L-shape adapter. The adapter was designed so that the measurement direction can be aligned with the vertical direction of the shaker. A rigid rectangular prism with a known mass was fixed onto the force sensor. The acceleration of the adapter and the acceleration of the prism were measured by two single-axial accelerometers. The dynamic inertia force of the prism was recorded by the force sensor. The inertia force of the prism was also calculated as the product of the acceleration measurement and the mass of the prism. This information was utilized as the baseline for the calibration of the force sensor. For the estimation of the frequency response function, the calibration was carried out from 20 to 1000 Hz with an increment of 10 Hz. An average of 20 cycles of oscillation was performed at each frequency level.

In Feng et al. [24] and Feng [23], the ASB Lab piezoelectric force sensor was calibrated using oscillating forces. A piezoelectric actuator was utilized to generate the oscillating forces, Fig. 26.5. The amplitude and the frequency of the
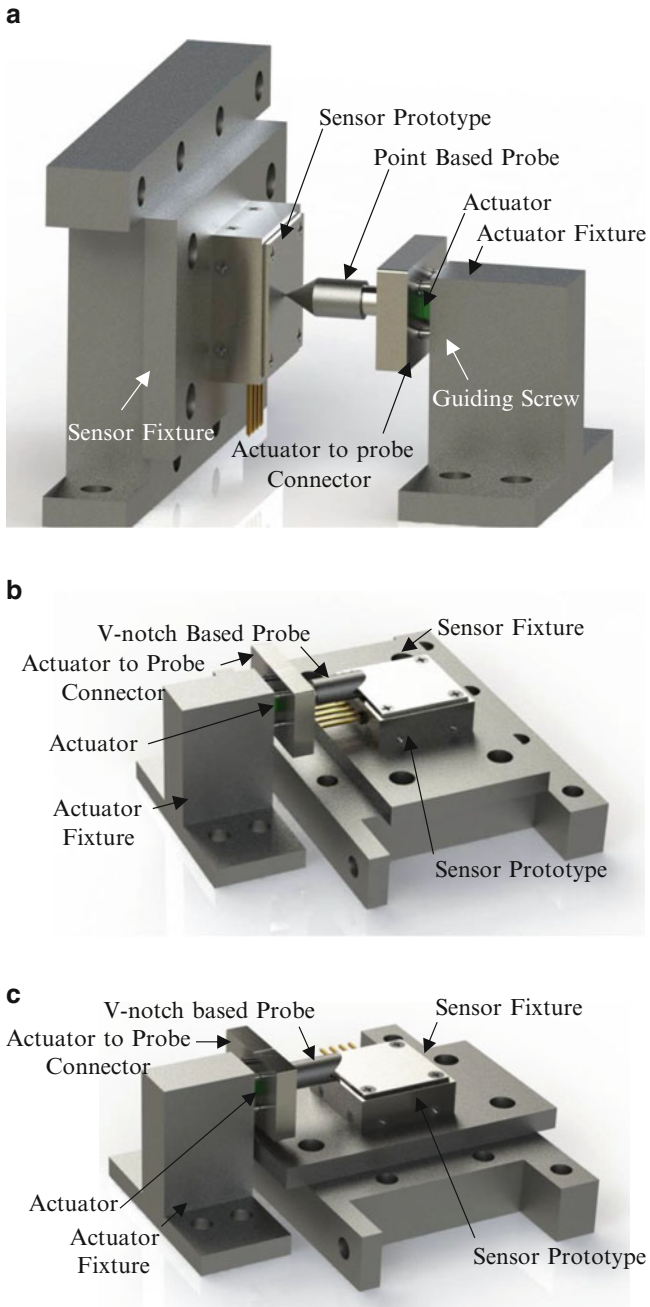
Fig. 26.5 Oscillating force calibration setup used in Feng et al. [24] and Feng [23] using a piezoelectric actuator: (**a**) for *x*, (**b**) *y,* and (**c**) *z* directions of the force sensor (Source: Feng, Y. & Nejat, G. (2015))
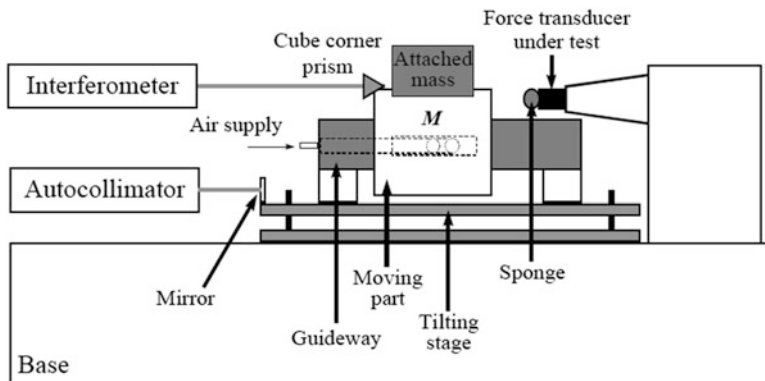
Fig. 26.6 Impulse force calibration setup in Fujii and Fujimoto [28] (Source: Fujii, Y., & Fujimoto, H. (1999), © IOP Publishing)

oscillating forces were varied by controlling the driving voltage of the actuator. The relationship between the amplitude of the driving voltage and the amplitude of the oscillating force was determined by conducting a pre-calibration test, in which a reference force sensor was used to measure the amplitude of the oscillating force while the amplitude of the driving voltage was also monitored. To calibrate all three components of the force sensor, a fixture was designed which aligned each axial direction of the force sensor with the actuation direction of the actuator, Fig. 26.5a–c. For each force component, the frequency of the input force was varied from 1 to 9 kHz with an increment of 1 kHz. For each frequency level, input forces with amplitude of 10 mN to 10 N were applied. The input force from the actuator was estimated according to the measured driving voltage as the baseline to compare with the force sensor measurement for the calibration of the force sensor.

3.2 Dynamic Calibration Methods Using Impulse Forces

In Fujii and Fujimoto [28], a calibration method for a one-axis force sensor using impulse forces was proposed. For the calibration of the force sensor, a rigid object with known mass was mounted on a pneumatic guideway to minimize the friction during the movement of the object, Fig. 26.6. The guideway and the force sensor under calibration were affixed to a stationary base. A rubber sponge was placed on the sensor to provide damping. The object was struck by an impact hammer and moved towards the sensor. Upon collision with the rubber damper, the object rebounded back. The locations of the object before and after the collision were recorded by an interferometer, while the impact force at the damper was measured by the force sensor. The difference in the velocity of the object due to the collision was determined. In order to calibrate the force sensor, the change in momentum of
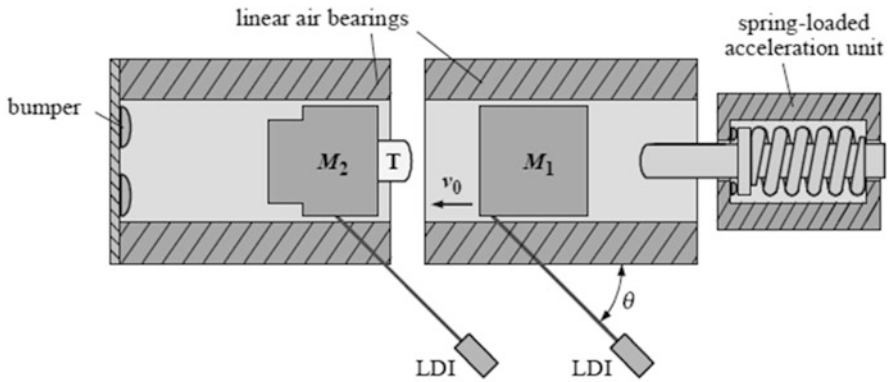
Fig. 26.7 Impulse force calibration setup in Bruns and Kobusch [29] with two moving objects. T represents the force sensor under calibration (Source: Burns, T. & Kobusch, M. (2001))

the object was determined following the impulse-momentum theorem and was used as the baseline for the calibration to compare with the change in momentum of the object determined from the time integration of the impact force measurements from the force sensor.

In Bruns and Kobusch [29], a calibration setup similar to that presented in Fujii and Fujimoto [28] was independently developed and presented for a one-axis force sensor using impact forces. The main difference was that instead of being fixed to the guideway, the sensor under calibration was placed onto a second rigid object on the guideway, Fig. 26.7. The first object was accelerated by a spring-loaded bolt and collided with the second object on which the force sensor was affixed. Interferometers were used to measure the displacements of the two objects. The velocities of the objects were determined in order to extrapolate the accelerations of the objects for estimating impact forces. The force measurements from the force sensor were then compared to force estimations in order to calibrate the sensor.

In Farm [30], a calibration approach based on the split Hopkinson pressure bar technique was proposed for a one-axis force sensor. The split Hopkinson pressure bar technique with a force sensor is depicted in Fig. 26.8a. In Farm [30], the sensor under calibration was fixed between an incident bar and a transmitted bar. Both bars had the same diameter and were made of the same material. A pre-calibrated strain gauge was attached to each bar to measure the strain experienced during the transmission of the wave. To predetermine the length of the wave, a pulse shaper bar with a specified length was attached in front of the incident bar. By striking the pulse shaper bar with a steel hammer, a wave could be sent to the incident bar and transferred through the force sensor to the transmitted bar. At the end of the transmitted bar, a shock absorber was affixed. The length of the pulse shaper bar was selected to be much smaller than that of the incident bar so that the impendence of the pulse shaper bar was negligible. The forces experienced at the force sensor were interpolated by using the strain measurements according to the one-dimensional wave propagation theory. The system was assumed to be under
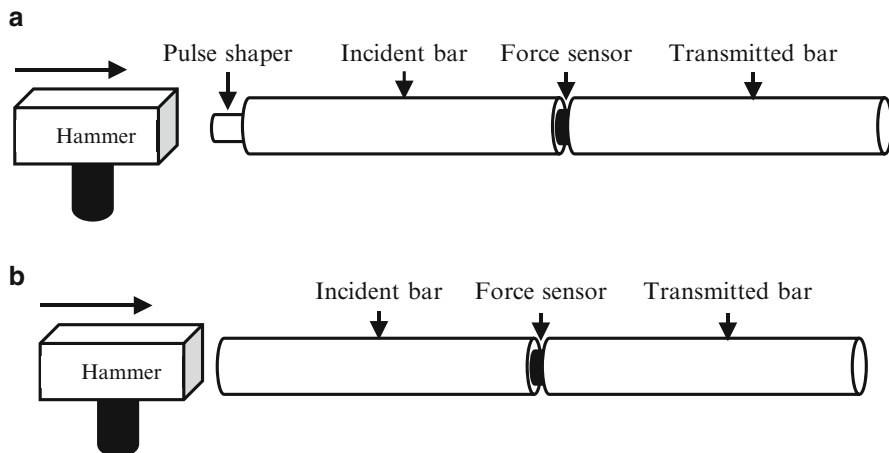
Fig. 26.8 Split Hopkinson pressure bar: (**a**) with a pulse shaper bar and (**b**) without a pulse shaper bar

quasi-static equilibrium, therefore, the strain developed in the transmitted bar was considered to be equal to the total strain developed in both the incident bar and the force sensor. The force measurements from the force sensor were compared with the force estimation from the strain measurements in order to calibrate the force sensor.

In Van Nuffel et al. [31], a calibration setup similar to that presented in Farm [30] was proposed without the pulse shaper bar. The wave was induced by striking the incident bar directly with a steel hammer. The split Hopkinson pressure bar technique without a pulse shaper bar is depicted in Fig. 26.8b. The forces generated were pulse based rather than plateau based as was generated when using the pulse shaper bar in Farm [30]. The impact forces were also estimated according to the one-dimensional wave propagation theory. The force measurements from the force sensor under calibration were compared with the estimated forces.

3.3 Dynamic Calibration Methods Using Step Forces

In Fujii [25], a calibration method using step forces was implemented for a one-axis force sensor. In this method, a rigid object with known mass was suspended over the force sensor under calibration by a stainless steel wire, Fig. 26.9. The object was fixed onto a vertical pneumatic linear bearing guideway to ensure vertical motion. The suspension wire was cut, and the object was able to free fall towards the force sensor. During the collision between the object and the sensor, an interferometer was used to record the displacement of the object. The acceleration of the object was calculated from the displacement measurements. The step force during the collision was estimated according to Newton's second law. The measurement from the force sensor was used to compare with the force estimation.
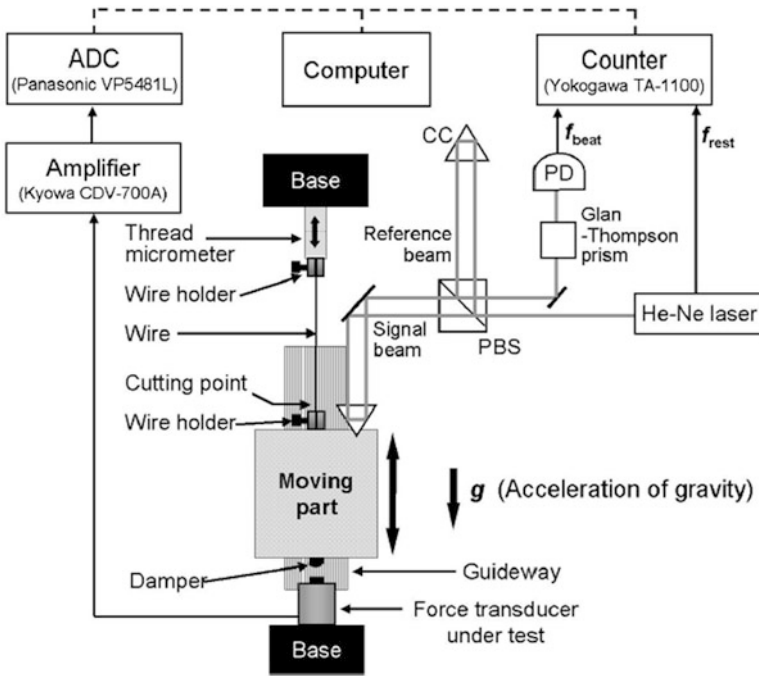
Fig. 26.9   Step force calibration setup in Fujii [25] (Source: Fujii, Y. (2003), © IOP Publishing)

4   Implementation of Piezoelectric Force Sensors in Micro/Meso-Milling Applications

The integration of piezoelectric force sensors for the measurement of cutting forces during micro/meso-milling processes can be utilized for the investigation of: (1) the effect of the feed rate on the cutting forces, and (2) real-time tool wear, deflection, and failure detection. Such applications of piezoelectric force sensors are discussed in this section.

4.1   Investigating the Effect of Feed Rate on Cutting Forces

In Dhanorker and Ozel [2], the Kistler 9265 piezoelectric dynamometer was integrated with a three-axis vertical milling CNC machine to investigate the effect of feed rate on the cutting force generation during meso-milling operations. The sensor was placed under the workpiece made of AL 2024-T6 aluminium alloy and two flat-end cutting tools with diameters of 1.5875 and 3.175 mm were used to cut slots on the workpiece. The cuts were made with a constant axial depth of 1.27 mm

and a constant spindle speed of 6000 rpm. For each tool, the normal forces and the feed forces were measured by the piezoelectric dynamometer at the feed rate of 0.265–4 μm per tooth. Using the same cutting tool, cutting forces increased when feed rate increased, however, the specific cutting forces (i.e., cutting forces required to cut a 1 mm^2 chip with a thickness of 1 mm) increased with a decreasing feed rate. It was also found that both the feed and normal forces increased with a decrease of the tool diameter under the same feed rate, which was due to the reduced stiffness of the cutter and increased elastic deformation caused by the reduction of the tool size.

4.2 Monitoring Tool Wear, Deflection, and Failure

In the development of a micro-machining center, the Kistler 9017 piezoelectric force sensor was utilized for tool wear monitoring during the machining process [15]. The micro-machining center platform design was based on a conventional column-and-knee-type machine. Since a high-speed spindle (up to 2.67 kHz) was required for a sufficient volumetric material removal rate, the operational frequency bandwidth of the force sensor (1 kHz), was insufficient. To address this issue, a Kalman filter (KF) technique using both force sensor measurements and acceleration measurements was implemented. The force sensor was placed between the workpiece and the base plate located on the micro-machining center platform. The accelerometer, Kistler 8778A500, was attached to the workpiece to measure vibration signatures. An impact hammer test was conducted to experimentally obtain the frequency response functions (FRFs) of the two sensors using a nonlinear fitting method. For measurements obtained by each sensor, a KF FRF was constructed and combined with the aforementioned FRFs to extend the operational frequency bandwidth of the force sensor. To verify the performance of the force sensor compensated with KF, experiments with spindle speeds of 5000, 40,000, and 80,000 rpm were conducted with the micro-machining machine. A two-fluted end mill cutting tool made of fine-grain tungsten carbide with a diameter of 200 μm was used. The material of the workpiece was aluminium 7075. The magnitude of the cutting forces was assumed to be the same regardless of the spindle speeds. The magnitude of the cutting forces measured at the spindle speed of 5000 rpm was used as a reference for comparison. For experiments with the spindle speeds of 40,000 and 80,000 rpm, the magnitude of the cutting forces obtained with and without the KF compensation were recorded and compared to the reference magnitude. The compensated measurements showed significant decrease in deviation from the reference force measurements when compared to the uncompensated measurements. The experiments verified that the KF technique was able to extend the frequency bandwidth of the force sensor, up to 2.67 kHz, in order to measure cutting forces.

In Garzon et al. [18], the HFD by Kistler was implemented on a combined machining platform. The platform was built and optimized on a Sarix SX 200 machine integrated with a high-speed ultra-precise spindle. The machine had

3-D micro-electrical discharge machining (EDM) capabilities for $4 + 2$ axis milling. Micro-milling was achieved with the integrated spindle which allowed for spindle speeds up to 160,000 rpm. A physical vapor deposition (PVD) coated two-edge cutting tool with a diameter of 200 μm was utilized for the cutting process. The material of the workpiece was steel 1.2343-56HRC. The HFD was fixed to the machining platform directly under the workpiece. Cutting tests with spindle speeds of up to 110,000 rpm were performed. The cutting forces during the engagement of each cutting-edge differed due to the run-out deviation of the cutting tool as well as the asymmetry between the cutting-edges. During the milling process, different cutting-edges of the milling tool were identified from the cutting force measurements of the dynamometer. In general, the dynamometer is capable of providing cutting force measurements for high-speed milling operations for the monitoring of tool wear and identification of the partial breakage of a cutting-edge.

In Kang et al. [16], the Kistler 9251 force sensor was used on a high-speed machining center, Makino V55, for measuring cutting forces for tool monitoring. The machining center was equipped with an air-turbine spindle. A 2-flute tungsten carbide micro end-milling tool of 200 μm diameter was used to perform full immersion slot milling. The workpiece was $20 \times 20 \times 10$ mm in size and made of Al6061-T6. The dynamometer was fixed onto the machining center directly under the workpiece. To verify the performance of the force sensor, milling experiments with spindle speeds of up to 66,000 rpm were conducted. Less than a 10 % error in cutting force measurements was achieved by the force sensor during the milling operations. The force sensor can be used to effectively monitor milling processes for tool wear, deflection, and failure.

5   Discussion and Conclusion

Miniature components are needed for a wide variety of devices and systems. Accurate measurement of cutting forces is important for understanding the cutting process during the micro/meso-milling operations in order to machine these components. Piezoelectric force sensors are effective sensors that can be used to measure cutting forces during micro/meso-milling applications.

Piezoelectric force sensors can be categorized into two groups based on the material used for their sensing elements which include either quartz crystals or ceramics. Quartz crystals are popular in commercially available force sensors due to their high mechanical strength and low hysteresis behavior. On the other hand, piezoelectric ceramic materials are easier and cheaper to mass produce and have a much higher piezoelectric sensitivity than quartz crystals. Table 26.1 provided an overview of the existing piezoelectric force sensors applicable for micro/meso-milling applications. The resolution of quartz-based force sensors can be high ranging from 2 to 10 mN, their maximum measurable force can also range from 10 N to 15 kN, and their maximum operational frequency bandwidth can be up to 5.1 kHz. Kalman filters can be designed to increase the upper limit of the frequency

bandwidth of such force sensors given the fixed natural frequency of the system. For example, in Chae and Park [15], the upper limit was increased by three times. Furthermore, by utilizing different sensing element materials and structures, the upper limit of the frequency bandwidth can also be increased. For example, as investigated in Feng et al. [24] and Feng [23], an upper frequency limit of 9 kHz was achieved with a square-shaped piezoelectric ceramic material as the sensing elements.

An industry standard does not exist for the dynamic calibration of piezoelectric force sensors, however, calibration approaches either utilize oscillating, impulse or step input forces. The actuation methods to provide these input forces vary between the different types of dynamic calibration techniques. For example, oscillating forces can be generated by exciting weighted masses with shakers or using piezoelectric actuators. Impulse forces can be generated by striking weighted masses with, for example, impact hammers and spring-loaded bolts, and step forces can be generated by gravitational forces using suspending weighted masses over the force sensors.

Development and integration of piezoelectric force sensors with high resolutions and natural frequencies is an open research area for the micro/meso milling of miniature 3-D components that are in high demand. Experimentation has verified the feasibility of integrating these types of sensors for high-speed micro/milling-machining. Future work also consists of online long-term utilization of such sensors for measuring cutting forces in a variety of different milling operations, for: (1) parameter optimization and feedback control and (2) tool wear monitoring.

References

1. Dow TA, Scattergood RO (2003) Mesoscale and microscale manufacturing processes: challenges for materials, fabrication and metrology. In: Proceedings of the ASPE winter topical meeting, vol 28, pp 1–19
2. Dhanorker A, Ozel T (2008) Meso/micro scale milling for micro-manufacturing. Int J Mech Manuf Syst 1(1):23–42
3. Ehmann KF, Bourell D, Culpepper ML et al (2005) International assessment of research and development in micromanufacturing. World Technology Evaluation Center, Baltimore, MD
4. Li H, Lai X, Li C et al (2008) Development of meso-scale milling machine tool and its performance analysis. Front Mech Eng China 3(1):59–65
5. Chae J, Park SS, Freiheit T (2006) Investigation of micro-cutting operations. Int J Mach Tool Manuf 46(3):313–332
6. Albrecht A, Park SS, Altintas Y et al (2005) High frequency bandwidth cutting force measurement in milling using capacitance displacement sensors. Int J Mach Tool Manuf 45(9):993–1008
7. Merchant ME (1945) Mechanics of the metal cutting process: II. Plasticity conditions in orthogonal cutting. J Appl Phys 16(6):318–324

8. Vogler MP, Liu X, Kapoor SG et al (2002) Development of meso-scale machine tool (mMT) systems. Technical papers-Society of Manufacturing Engineers-all series
9. Bissacco G, Gietzelt T, Hansen HN (2008) Force analysis in micro milling Al 6082 T6 in various engagement conditions. In: Proceedings of 4M 2008 conference multi-material micro manufacture, Whittles Publishing, Cardiff
10. Uhlmann E, Piltz S, Schauer K (2005) Micro milling of sintered tungsten–copper composite materials. J Mater Process Technol 167(2):402–407
11. Byrne G, Dornfeld D, Inasaki I et al (1995) Tool condition monitoring (TCM)—the status of research and industrial application. CIRP Ann Manuf Technol 44(2):541–567
12. Gautschi G (2002) Piezoelectric sensorics: Force, strain, pressure, acceleration and acoustic emission sensors, materials and amplifiers. Springer, New York
13. Kistler Group (2009) 3-component dynamometer, Type 9265B. Switzerland
14. Kistler Group (2005a) 3-component force sensor, Type 9016B4, 9017B, 9018B. Switzerland
15. Chae J, Park SS (2007) High frequency bandwidth measurements of micro cutting forces. Int J Mach Tool Manuf 47(9):1433–1441
16. Kang IS, Kim JH, Hong C et al (2010) Development and evaluation of tool dynamometer for measuring high frequency cutting forces in micro milling. Int J Precis Eng Manuf 11(6):817–821
17. Kistler Group (2005b) 3-component force sensor, Type 9251A, 9252A, 9250A4, 9251A4. Switzerland
18. Garzon M, Adams O, Veselovac D et al (2012) High speed micro machining processes analysis for the precision manufacturing. Proc CIRP 1:609–614
19. Kistler Group (2005c) MiniDyn, Type 9256C. Switzerland
20. Transchel R, Stirnimann J, Blattner M et al (2012) Effective dynamometer for measuring high dynamic process force signals in micro machining operations. Proc CIRP 1:558–562
21. Mekid S (2014) Design and testing of a micro-dynamometer for desktop micro-milling machine. Adv Mater Res 902:267–273
22. Feng Y, Shao F, Nejat G (2014). Prototype development of a three-component force sensor for meso milling applications. In: CIRP sponsored international conference on virtual machining process technology
23. Feng Y (2015) Development and calibration of a 3-component force sensor for meso-milling. Master's Thesis, University of Toronto
24. Feng Y, Xu Y, Nejat G (2015) Dynamic calibration of a three-component force sensor for meso-milling applications. In: CIRP sponsored international conference on virtual machining process technology
25. Fujii Y (2003) Proposal for a step response evaluation method for force transducers. Meas Sci Technol 14(10):1741
26. Kumme R (1998) Investigation of the comparison method for the dynamic calibration of force transducers. Measurement 23(4):239–245
27. Park YK, Kumme R, Kang DI (2002) Dynamic investigation of a three-component force-moment sensor. Meas Sci Technol 13(5):654
28. Fujii Y, Fujimoto H (1999) Proposal for an impulse response evaluation method for force transducers. Meas Sci Technol 10(4):31
29. Bruns T, Kobusch M (2001) Impulse force calibration: design and simulation of a new calibration device. In: Proceedings of the 17th international conference on force, mass, torque and pressure measurements IMEKO TC3, pp 17–21
30. Farm J (2003) Split Hopkinson pressure bar technique for dynamic calibration of force transducers. SP Swedish National Testing and Research Institute, Boras, Sweden
31. Van Nuffel D, Peirs J, De Baere I et al (2012) Calibration of dynamic piezoelectric force transducers using the Hopkinson bar technique. In: 15th international conference on experimental mechanics. INEGI-Instituto de EngenhariaMecânica e Gestão Industrial. ICEM15-2012, pp 341–342

Chapter 27
Magnetically Driven Microrobotics for Micromanipulation and Biomedical Applications

Xiaodong Zhang and Mir Behrad Khamesee

Abstract Magnetic navigation is a promising technology for micromanipulation. This chapter firstly provides a brief review of magnetically driven microrobots developed for micromanipulation and biomedical applications such as medical surgery and drug delivery. Then a magnetically driven microrobotics system is introduced in detail to explain the procedure of developing a magnetic levitation stage. In order to achieve the persistent navigation of the microrobot in various environments, this chapter proposes a sensor switching mechanism that combines magnetic flux measurement based position determination and optical sensor based position detection. In addition, the concept of minimum magnetic potential energy point is used to predict the magnetic force on the microrobot when there is physical contact between the magnetized object and environments. As a further exploration of magnetic force determination mechanism, this chapter also describes the concept of remotely controlled drug delivery with contact force perception at the operator side.

Keywords Microrobotics • Magnetic levitation • Micromanipulation • Electro-magnetic actuators • Magnetic field • Non-contact manipulation • Magnetic flux pattern • Hall-effect sensors • Feedforward control • Off-board force measurement • Force control • Biomedical application • Drug delivery • Capsule robot

1 Introduction

The magnetically driven technique has a long development history. Among different magnetically driven techniques, magnetic levitation is the most frequently studied and has made a great progress recently. In magnetic levitation, a system transfers the magnetic energy into the kinematic/potential energy of a moving part. The energy transformation is achieved remotely by means of magnetic field. Magnetic levitation has been conducted on materials that have different magnetic permeability relative

X. Zhang • M.B. Khamesee (✉)
University of Waterloo, 200 University Ave W, Waterloo, ON, Canada N2L 3G1
e-mail: x442zhan@uwaterloo.ca; khamesee@uwaterloo.ca
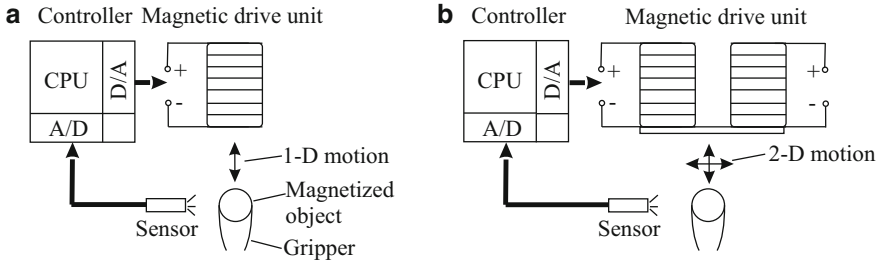
Fig. 27.1 (**a**) 1-D levitation using one electromagnet; (**b**) 2-D levitation using two electromagnets

to air, for instance, paramagnetic levitation [1, 2] of materials that have $\mu_r > 1$, superconducting levitation [3, 4] of materials that have $\mu_r = 0$, and diamagnetic levitation [5, 6] of materials that have $0 < \mu_r < 1$. Magnetic levitation systems were implemented in different industries, ranging from macro to micro, such as maglev trains [7], linear motors [8], wind tunnel testing [9], academic education [10], and cell cultivation [11].

A magnetic levitation system mainly consists of an energy source, a levitation robot, sensors, and a controller unit. Figure 27.1 shows the basic schematic of a magnetic levitation system for 1-D and 2-D levitation. The magnetic drive unit consisting of one or more electromagnets produces energy for levitation and manipulation. The levitation robot is a magnetized object. In order to save energy, the magnetized object is commonly made from ferromagnetic materials that have high magnetic permeability and from a permanent magnet that has strong remnant magnetic flux density. A gripper is installed on the magnetized object to perform special tasks. Earnshaw's theorem indicates that pure interaction between a magnetic energy source and a magnetized object cannot maintain a stable levitation. The feedback dynamics of the levitation robot are required. Therefore, position sensor and controller unit are necessary. A single electromagnet can be used for the levitation of an object and motion in the vertical direction (Fig. 27.1a). If multiple electromagnets are used, the vertical and the horizontal motions can be achieved simultaneously (Fig. 27.1b). The horizontal motion is achieved by changing the current ratio in each electromagnet.

The magnetic levitation microrobotics (MLM) is a new emerging research area that has promising potential applications in the fields of micro-electromechanical system (MEMS) fabrication, micro-accuracy actuation, and biomedical surgery. Based on the scale of end-effectors, MLM can be categorized into two types: millimeter scale magnetic levitation manipulators that have micro-scale navigation accuracy, and magnetic levitation microrobots that have micro-scale end-effectors. For millimeter scale manipulators, the end-effectors are larger than 1 mm. This type of manipulators is used for micromanipulation and micro-assembly. For micro-scale actuators, the sizes of the end-effectors are less than 1 mm. This type of microrobots is used in biomedical applications. As the size of an actuator decreases, the dominant force on the actuator changes from bulk force to surface tension force.

   According to their applications, we classify the MLMs into magnetic micromanipulators and magnetically navigated microrobots.

1.1 Magnetic Micromanipulators

Modern scientific progress promotes the development of micro-manipulators that are capable of working in delicate environments with high precision. Conventional solutions to micromanipulation involve designing lead-screw driving stages and piezoelectric actuators. There are many successful applications of these technologies [12–14]. However, lead-screw driving stages are characterized by stiction, backlash, a mechanical arm that introduce friction, and impulse force that can damage manipulated objects. The piezoelectric actuators have a high nonlinearity and hysteresis effect. Modeling and control of piezoelectric actuators are very complex [15, 16]. In contrast, there is no mechanical connection between end-effectors and the magnetic energy source in MLMs. Therefore, MLMs have the benefits of being contactless, having low noise, having no backlash, being frictionless, and offering multi degree of freedom motion. In addition, the advancement in stabilizing levitation makes the accuracy of magnetic navigation comparable to or even better than conventional arm robot in micromanipulation [17, 18]. Therefore, the MLM is becoming more and more popular in high accuracy micropositioning systems.
   Successful magnetic levitation micromanipulators have been reported in literature. Kim [19] proposed a planar stage for the potential application in photolithography. The stage used four linear permanent-magnet motor to support the levitator. A Halbach permanent magnet array was installed on the levitator to interact with the stator for six degree of freedom (DOF) motion. The translation accuracy of this stage was 5 nm, which is super accurate. In [20], a planar maglev positioning system for micromanipulation was presented. The stator of the system was made of six coils, three vertical coils for controlling the vertical levitation and three horizontal coils for controlling horizontal motion. Corresponding to six stationary coils, six permanent magnets were placed inside the coils to support the levitator. This stage has 6-DOF motion ability and 10 μm translation accuracy. The dust-free, high accuracy, and high speed properties of the magnetic micromanipulator bring improvement to micro-automation industries such as semiconductor manufacturing [21].

1.2 Magnetically Navigated Microrobots

Magnetic micronavigation uses a magnetic field to remotely navigate a microrobot. Thanks to their contactless property, magnetically navigated microrobots (MNM) are appropriate for micromanipulations in confined and human-unaccessible environments. Among these, biomedical applications such as microsurgery and deliv-

ering drugs to a special destination inside the human body are becoming more and more popular. For example, microsurgery can be done by sending an MNM directly to the destination of organ through vessels. A magnetized gastrointestinal endoscope can be sent directly to the human stomach without using a long soft pipe. Many studies on MNM were reported in literature [22–25].

An eye surgery micromanipulation system was presented in [26]. The system consists of six iron-core electromagnets and a small surgical robot. The size of the robot was 2 mm in length and 1.5 mm in diameter. The robot was guided under a digital camera. A maximum of 29 μm positioning deviation was reported in the study. In [27], delivering an endoscope capsule robot to stomach was reported. In this study, the capsule robot integrated a video camera, permanent magnets, and image emission circuit. A Niobe magnetic navigation system developed by stereo-taxis was used as the navigation energy source. The navigation was fulfilled under fluoroscopic images. A variety of platforms have been developed for micro/nano scale particle tracking [28], cell and molecule manipulation [29, 30]. One should be aware of the difficulty of localizing the navigated microrobot, specifically in nontransparent environments. Current localization methods include microscope, X-ray, and ultrasound imaging.

This chapter is organized as follows: Sect. 2 introduces the magnetic levitation system developed in the authors group. This includes principle of magnetic lev-itation, the modeling of the magnetic field in the workspace, and motion control of a microrobot with three degree of freedom. The sensor switching mechanism for navigating a microrobot in different environments is then presented in Sect. 3. A new force determination mechanism using magnetic flux measurement and its biomedical application is introduced in Sect. 4. Section 5 briefly summarizes this chapter.

2 Magnetically Driven Microrobotic System

2.1 Magnetic Field Source

Active navigation of a magnetized object requires a controlled magnetic field source. Generally, the magnetic field can be generated by current-controlled electromagnets [31, 32], permanent magnets held by a position-controlled manipulator [25, 27], and Helmholtz coils [24]. The use of a permanent magnet provides limited motion con-trol accuracy, since it is challenging to adjust the position of a manipulator actively at the relatively high frequency necessary for stable levitation of a magnetized object. The magnetic field and its gradient are independent in a Helmholtz coils system. A system that uses Helmholtz coils is limited to 3D translation of an object.

Electromagnets are commonly used for multi-dimensional manipulation of a magnetized object. The magnetic field generated by electromagnets is controlled by a current. The multi-degree of motion freedom of an object is achieved by setting up

special configurations of multi-electromagnets. Since the strength of the magnetic field decays rapidly as the distance from electromagnet increases, an iron-core is commonly used to significantly enhance the magnetic field strength outside the electromagnet. This enhanced magnetic field can then expand the workspace of the magnetized object.

2.2 Magnetic Levitation Force

In a contact-free levitation, a magnetic levitation system generates satisfied levitation force to compensate the gravity force on a magnetized object, and to maintain a desired dynamics of the object. A small object that is evenly magnetized with a total magnetization M (unit A/m) is placed in an external magnetic field, the total magnetic potential energy on the object is

$$U = -\mathbf{M} \cdot \mathbf{B} \tag{27.1}$$

Using the virtual displacement method, the magnetic force on the small magnetized object by gradient magnetic field can be expressed as

$$F = \nabla(\mathbf{M} \cdot \mathbf{B}) \tag{27.2}$$

In addition to the magnetic force, the magnetic field aligns the magnetization direction of the object to the direction of the magnetic field by a magnetic torque:

$$\boldsymbol{\tau} = \mathbf{M} \times \mathbf{B} \tag{27.3}$$

where $\mathbf{B}$ is the magnetic flux density of the magnetic field where the object stays. Its unit is Tesla. If we could assume that the magnetization direction of the object is in z-direction [22], i.e., $\mathbf{M} = [0, 0, M]$, then Eq. (27.1) can be expanded as

$$F_x = MV \frac{dB_z}{dx} \tag{27.4}$$

$$F_y = MV \frac{dB_z}{dy} \tag{27.5}$$

$$F_z = MV \frac{dB_z}{dz} \tag{27.6}$$

The force models indicate that the forces on the small magnetized object are related to the gradient of the magnetic field at the levitation location and the magnitude of the magnetization of the object. This can be used to guide the development of a magnetic levitation system, i.e., fabricate a platform that generates high magnetic field gradient and select object that has strong magnetization to

save energy. In this chapter, the three-dimensional navigation of a small magnetized object is presented. Orientation of the object is not considered here. In order to control the orientation of the object, extra electromagnet setup is required.

2.3 Magnetic Levitation Stage

Equations (27.4)–(27.6) indicate that a non-uniform external magnetic field is required to impose magnetic forces on a magnetized object. In the non-uniform magnetic field, there is a point that has the minimum magnetic potential energy. Khamesee [33] explained that the magnetized object tends to move toward the minimum magnetic potential energy point in the external magnetic field. This point, also known as the point that has the maximum magnetic flux density, is defined as B_{max} point. A single electromagnet can generate a single B_{max} point on a horizontal plane below the electromagnet, and can produce magnetic gradient force to compensate the gravity force. However, the horizontal location of the B_{max} point cannot be changed. Therefore, only one-directional levitation is achievable using a single electromagnet.

In order to obtain the multi-dimensional levitation of a magnetized object, a combination of electromagnets is required. However, multiple electromagnets produce multiple B_{max} points, which make the levitation of a single magnetized object uncontrollable. Khamesee [34] proposed using an iron pole-piece to connect multiple electromagnets. The induced magnetization of the pole-piece produces a single B_{max} point below the electromagnets.

In the developed magnetic levitation system shown in Fig. 27.2, six identical electromagnets powered with current are the magnetic field source. A disc shape
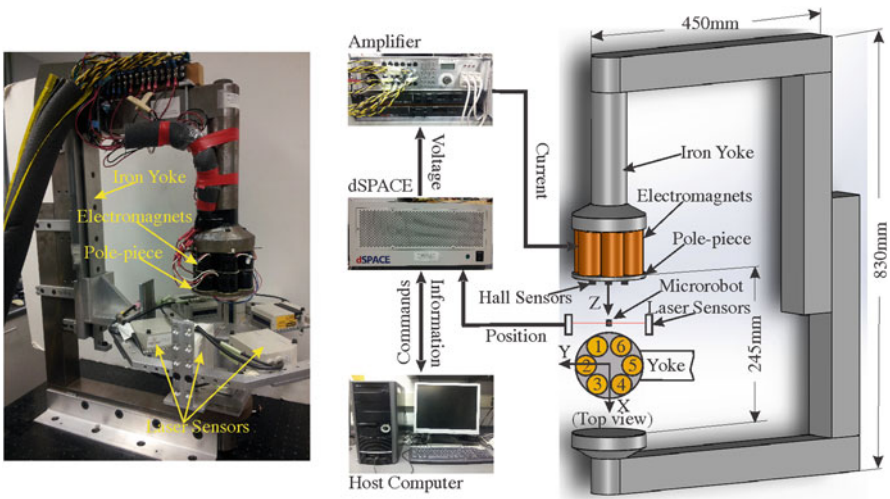
Fig. 27.2 Configuration of magnetic levitation system

	Evenly Loaded $I_1 = I_2 = I_3 = I_4 = I_5 = I_6$	Unevenly Loaded $I_2 > I_{1,3} > I_{4,6} > I_5$
Without Yoke	**a**	**b**
With Yoke	**c**	**d**
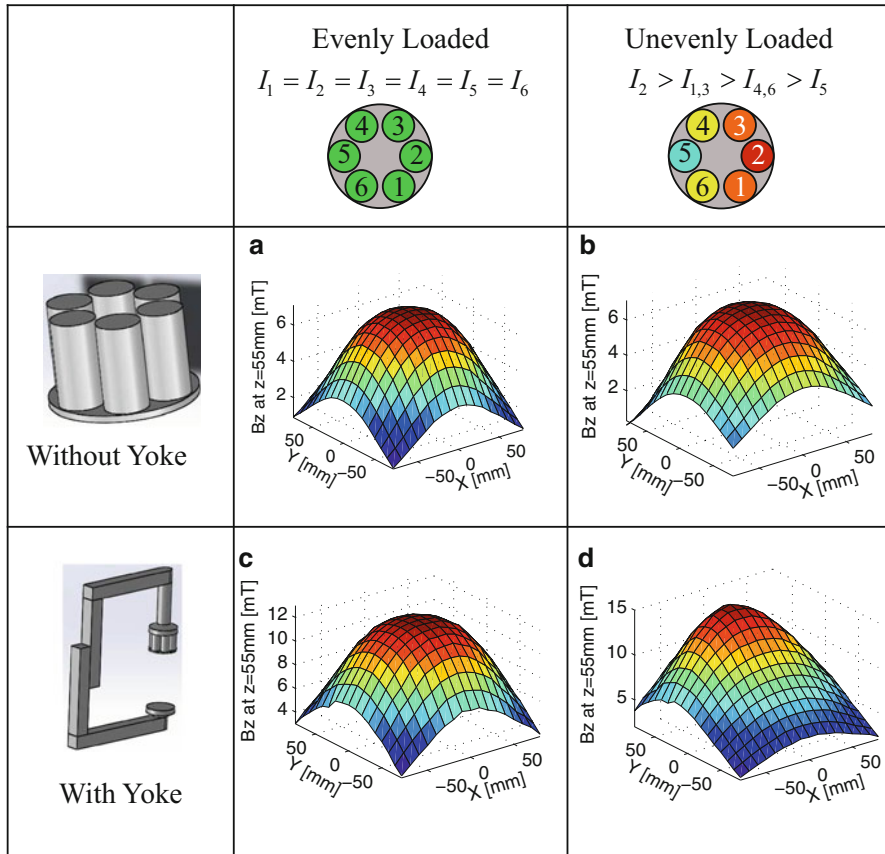
Fig. 27.3 (**a**) Magnetic flux density without yoke while all electromagnets are evenly loaded; (**b**) Magnetic flux density without yoke while electromagnets are unevenly loaded; (**c**) Magnetic flux density with yoke while all electromagnets are evenly loaded; (**d**) Magnetic flux density with yoke while electromagnets are unevenly loaded

soft iron pole-piece connects the electromagnets and configures the magnetic field in the workspace. A soft iron yoke is installed to generate a closed loop magnetic circuit and increase the magnetic field strength in the workspace. Simulation results in Fig. 27.3 are presented to show the concept of changing the B_{max} point location and the effect of the iron yoke on increasing the intensity of the magnetic field at $z = 55$ mm below the pole-piece.

As shown in Fig. 27.3a, b, the location of the maximum magnetic flux density point in the workspace is located in the center of a horizontal plane if the electromagnets are equally loaded. Otherwise, the maximum magnetic flux density point moves toward the electromagnet that is loaded with more current. Figure 27.3c, d shows that when an iron yoke is installed, the magnetic field intensity and the horizontal gradient of the magnetic field are increased significantly. The soft iron
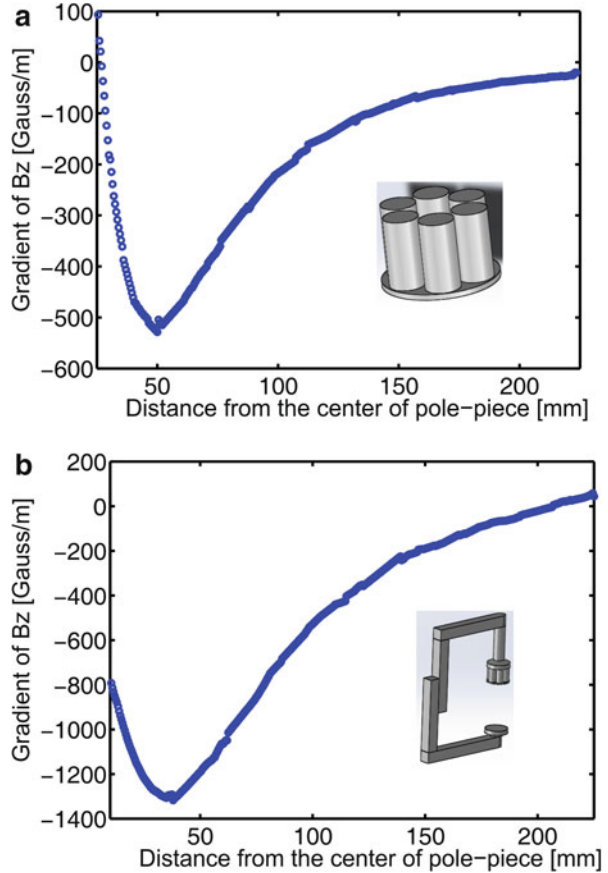
yoke also increases the vertical gradient of the magnetic field in the workspace, which in turn increases the levitation force on the levitated object. Figure 27.4 shows the vertical gradient of the magnetic field below the pole-piece. It shows that with the iron yoke, the gradient of the magnetic field is twice as high as the gradient without an iron yoke.

The magnetic flux density in the air-gap of the levitation stage while electromagnets are equally loaded with 1 A current is shown in Fig. 27.5. On planes close to the pole-piece, the B_{max} point is not unique. Stable levitation of a single magnetized object on these planes is not achievable. On planes relatively far from the pole-piece, only one B_{max} point exists. Stable levitation of a single magnetized object on these planes is achievable. The workspace of the developed magnetic levitation system is 65–95 mm below the pole-piece. In this section, the gradient of the magnetic field is linear to the distance from the pole-piece.
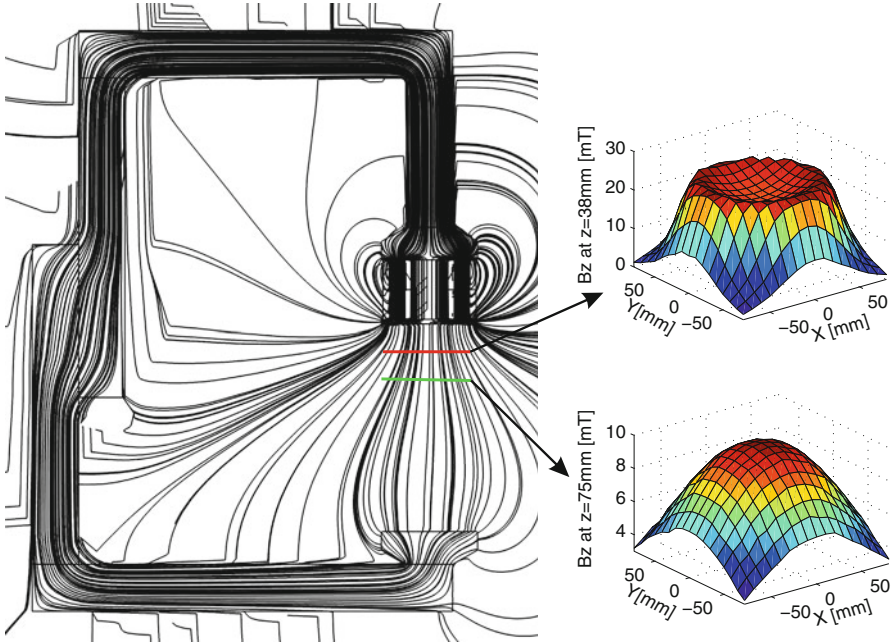
Fig. 27.5 Qualitative analysis of the magnetic field in the workspace

2.4   Modeling the Magnetic Field in the Workspace

Modeling the magnetic field in the workspace is crucial to developing the dynamic mode and to the motion control of the levitated object. For the developed magnetic levitation system, it is impossible to find a closed form for modeling the magnetic field in the air-gap, because the soft-iron yoke, iron-core electromagnets, and pole-piece simply result in high nonlinearity in the whole system. Therefore, an experimental method is applied. A Hall-effect sensor was used to scan the workspace and to record the magnetic flux density in real time. Figure 27.6 illustrates the magnetic flux density measured at 75 mm below the pole-piece while electromagnet 1 was powered with 1 A current.

The magnetic field produced by one electromagnet is linear to the current in this electromagnet. According to the measured data presented in Fig. 27.6, the magnetic field produced by electromagnet 1 can be modeled as

$$B_{z1} = (a_{1x}x^2 + b_{1x}x + c_{1x})I_1 \qquad (27.7)$$

where a_{1x} is a function of the z. b_{1x} and c_{1x} are functions of z and y. I_1 is the current in electromagnet 1.

The magnetic flux density produced by other electromagnets can be modeled in the same way. The total magnetic flux density in the workspace is the linear summation of magnetic flux density produced by all electromagnets:

Fig. 27.6 Magnetic flux density measured at 75 mm below the pole-piece while electromagnet 1 is powered with 1 A current
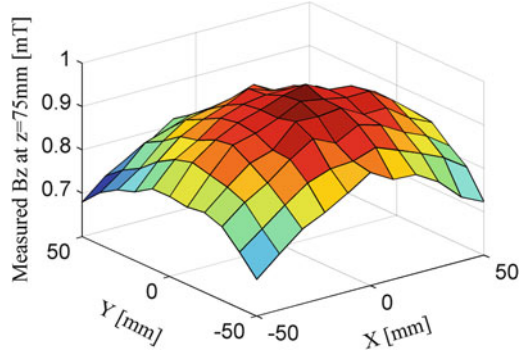
$$B_z = \sum_{i=1}^{6} B_{zi} \tag{27.8}$$

By considering the geometric symmetry, the total magnetic flux density can be expressed as

$$
\begin{aligned}
B_z = {} & (a_{1x}x^2 + b_{1x}x + c_{1x})I_1 + (a_{2x}x^2 + c_{2x})I_2 \\
& + (a_{1x}x^2 - b_{1x}x + c_{1x})I_3 + (a_{4x}x^2 + b_{4x}x + c_{4x})I_4 \\
& + (a_{5x}x^2 + c_{5x})I_5 + (a_{4x}x^2 - b_{4x}x + c_{4x})I_6
\end{aligned} \tag{27.9}
$$

where a_{2x}, a_{4x}, and a_{5x} are functions of z. b_{4x}, c_{2x}, c_{cx}, and c_{5x} are functions of z and y. $I_i(i = 1, \ldots, 6)$ are the current in the number i electromagnet.

Substitute Eq. (27.9) into Eqs. (27.4)–(27.6), and linearize at the center $[0, 0, z_0, I_0]$ of the workspace. The magnetic forces on a magnetized object with volume magnetization M are expressed as

$$
\begin{aligned}
F_x = M\frac{\partial B_z}{\partial x} = {} & 2xI_0\,(a_{x1}z_0 + b_{x1}) \\
& + (a_{x2}z_0 + b_{x2})\,(i_1 - i_3 + i_4 - i_6)
\end{aligned} \tag{27.10}
$$

$$
\begin{aligned}
F_y = M\frac{\partial B_z}{\partial y} = {} & 2yI_0\,(a_{y1}z_0 + b_{y1}) \\
& + (a_{y2}z_0 + b_{y2})\,(i_1 + 2i_2 + i_3 - i_4 - 2i_5 - i_6)
\end{aligned} \tag{27.11}
$$

$$
\begin{aligned}
F_z = M\frac{\partial B_z}{\partial z} = {} & (a_z z + b_z)\,I_0 + a_z I_0\,(z - z_0) \\
& + (a_z z_0 + b_z)\,(i_i + i_2 + i_3 + i_4 + i_5 + i_6)
\end{aligned} \tag{27.12}
$$

I_0 is the current in electromagnets when the levitated robot stays at z_0. All coefficients in Eqs. (27.10)–(27.12) are deduced from coefficients in Eq. (27.9). $i_j, j = 1, \ldots, 6$ are the perturbed current in the jth electromagnet.

In a contact free levitation state, the dynamic of the levitated object is obtained using the Newton's Second Law:

$$m\ddot{x} = 2xI_0\left(a_{x1}z_0 + b_{x1}\right) + \left(a_{x2}z_0 + b_{x2}\right)\underbrace{\left(i_1 - i_3 + i_4 - i_6\right)}_{u_x} \tag{27.13}$$

$$m\ddot{y} = 2yI_0\left(a_{y1}z_0 + b_{y1}\right) + \left(a_{y2}z_0 + b_{y2}\right)\underbrace{\left(i_1 + 2i_2 + i_3 - i_4 - 2i_5 - i_6\right)}_{u_y} \tag{27.14}$$

$$m\ddot{z} = -a_z I_0\left(z - z_0\right) + \left(a_z z_0 + b_z\right)\underbrace{\left(i_i + i_2 + i_3 + i_4 + i_5 + i_6\right)}_{u_z} \tag{27.15}$$

where m is the total mass of the levitated object.

2.5   Three-Dimensional Navigation of a Microrobot

The dynamic models shown in Eqs. (27.13)–(27.15) indicate that the levitation in horizontal directions is internally stable, since the dynamic model has two poles on the left side of the imaginary axis. However, the vertical direction dynamic model has one pole on the right side of the imaginary axis. The levitation in vertical direction is internally unstable. Therefore, feedback control is necessary for a stable levitation. Using the derived dynamics model, a PID plus feed forward controller is proposed to control the three-dimensional motion of a small permanent magnet in a contact free space. The schematic of the controller is shown in Fig. 27.7. In the vertical direction levitation, the feed forward controller provides the major control signal; the PID controller is used to compensate for the actual signal.

Three virtual control inputs (u_x, u_y, and u_z) are converted to actual current input to electromagnets using the pseudo inverse method, which minimizes the normal of input signal. The real position of a levitation robot is detected using three sets of high accuracy laser-beam sensors.
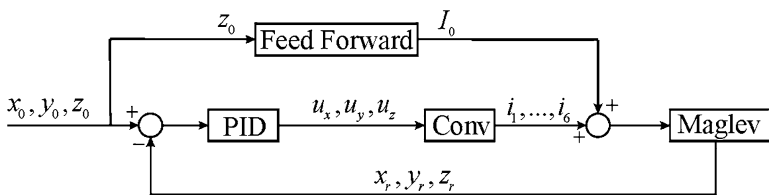
Fig. 27.7 Schematic of the PID plus feed forward controller for three-dimensional motion control
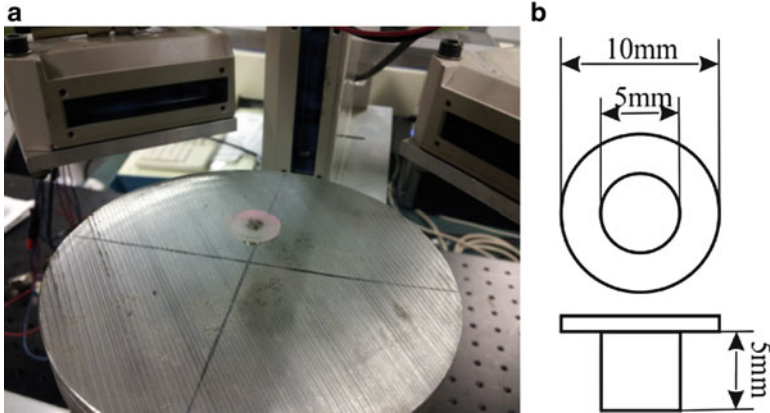
Fig. 27.8 (**a**) The microrobot is levitated; (**b**) Dimensions of the microrobot

Three-dimensional motion control experiment was conducted to validate the performance of the proposed controller. In the experiment, the levitated microrobot was a small cylindrical permanent magnet that had 5 mm in diameter and 5 mm in height. Its remnant magnetic flux density is 0.29 T. The configuration of the levitated microrobot is shown in Fig. 27.8.

The step motion response was recorded in the experiment. The experimental results are presented in Fig. 27.9. The experimental performance shows that the proposed PID plus feed forward controller can provide 20 μm motion accuracy in three axes. However, it should also be noted that cross-coupling effect exists among three motion axes. The motion in y direction has a significant impact on the stable levitation of the other two directions. In order to reduce the cross-coupling effect, an advanced controller, such as an LQG controller [35], can be implemented to control the motion of the microrobot.

3 Navigation of the Microrobot in a Non-transparent Environment

Active navigation of a magnetized microrobot depends on stabilizing the inherent unstable dynamics of the microrobot. It can be achieved by regulating the current in electromagnets based on the feedback of the microrobot position. Generally, precision laser-beam sensors are used to provide the feedback position of the microrobot for a motion control system. The laser beam sensors can provide sub-micrometers motion accuracy. However, the optical sensors only work in transparent work environments. There is a promising limitation when they are used to navigate a microrobot in a work environment with optical blockage. X-ray and ultrasound devices are commonly used for navigating a microrobot in a nontransparent
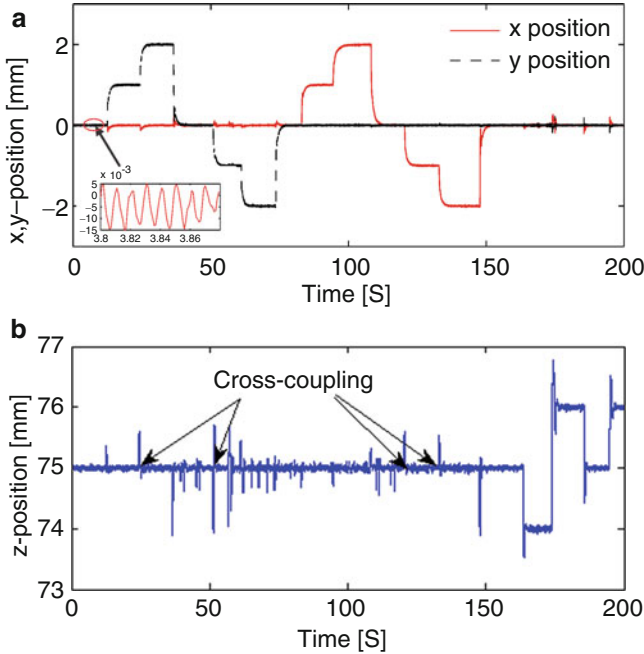
Fig. 27.9 (**a**) Motion trajectory in x- and y-direction; (**b**) Motion trajectory in z-direction

environment [27, 36]. However, the X-ray will be harmful to living cells, the ultrasound penetration depends on the environment, and the complexity of these systems also limits their application in real industry.

To address the limitation of optical blockage, the use of magnetic flux measurement is a substitution. The magnetic flux can penetrate any nonmagnetic materials without degrading its strength. Therefore, it is not limited by optical blockage. Mehrtash in [37] presented the one-directional navigation of a magnetized microrobot using magnetic flux measurement. Zhang in [38] presented the dual-axial motion control of a magnetically guided robot using four Hall-effect sensor. The fundamental is the relation between the location of the levitated microrobot and the distribution of the magnetic flux in the workspace. Therefore, by mapping this relation and regulating the magnetic field pattern in the workspace using magnetic flux feedback, the internally stable dynamics of a levitated microrobot can be controlled using an open-loop strategy in a small workspace.

Although there are advantages in using magnetic field measurement to determine the position of the microrobot, an open loop control of the vertical direction levitation of the Maglev system was internally unstable due to the presence of gravity force. The stable levitation in this direction requires the dynamic information of the microrobot. The magnetic levitation system in this chapter has a large air-gap. The vertical direction levitation cannot be maintained by magnetic field measurement-based position feedback controller. This is because the low resolution
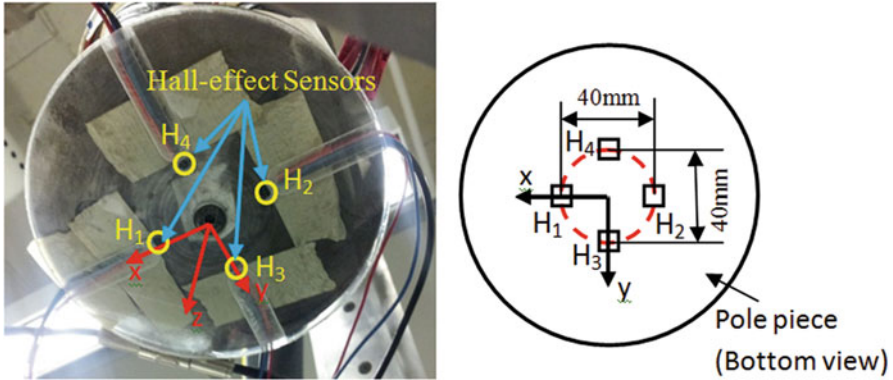
Fig. 27.10 Installation of Hall-effect sensors [38]

of Hall-effect sensors used in this study cannot detect the magnetic field of the microrobot owing to the interference produced by electromagnets.

3.1 Modeling Magnetic Field Pattern for Navigation

In this chapter, the navigation of the microrobot on a horizontal plane using the magnetic flux measurement based position determination is presented. As is shown in Fig. 27.10, four Hall-effect sensors are installed symmetrically at the bottom of the pole-piece (as pointed out in Fig. 27.2) to measure the magnetic flux pattern in the workspace. In the figure, H1, H2, H3, and H4 are four Hall-effect sensors. H1 and H2 are installed in the *x*-axis to measure the magnetic flux pattern in the *x*-axis. H3 and H4 are installed in the *y*-axis to measure the magnetic flux pattern in the *y*-axis. Output voltage difference of the two Hall-effect sensors is used to represent the magnetic flux in each axis. For example, V_{dx} is the magnetic flux in the *x*-axis, V_{dy} is the magnetic flux in the *y*-axis.

As discussed above, the microrobot will stay at the $B_{\max}$ point if the levitation is in a contact-free space. Therefore, the relation between magnetic field pattern and microrobot positions can be mapped by measuring the magnetic flux in the workspace when the microrobot is navigated using laser-beam sensors. Figure 27.11 shows the measured magnetic field while the microrobot is navigated using laser-beam sensors on the horizontal plane that locates at 78 mm below the pole-piece.

According to the measured magnetic flux pattern, the real positions of the microrobot in horizontal directions are determined by mapping the measured magnetic flux using high order polynomial equations. The inputs for both models were the output voltage differentials V_{dx} and V_{dy} of the Hall-effect sensors. In Eq. (27.16), the
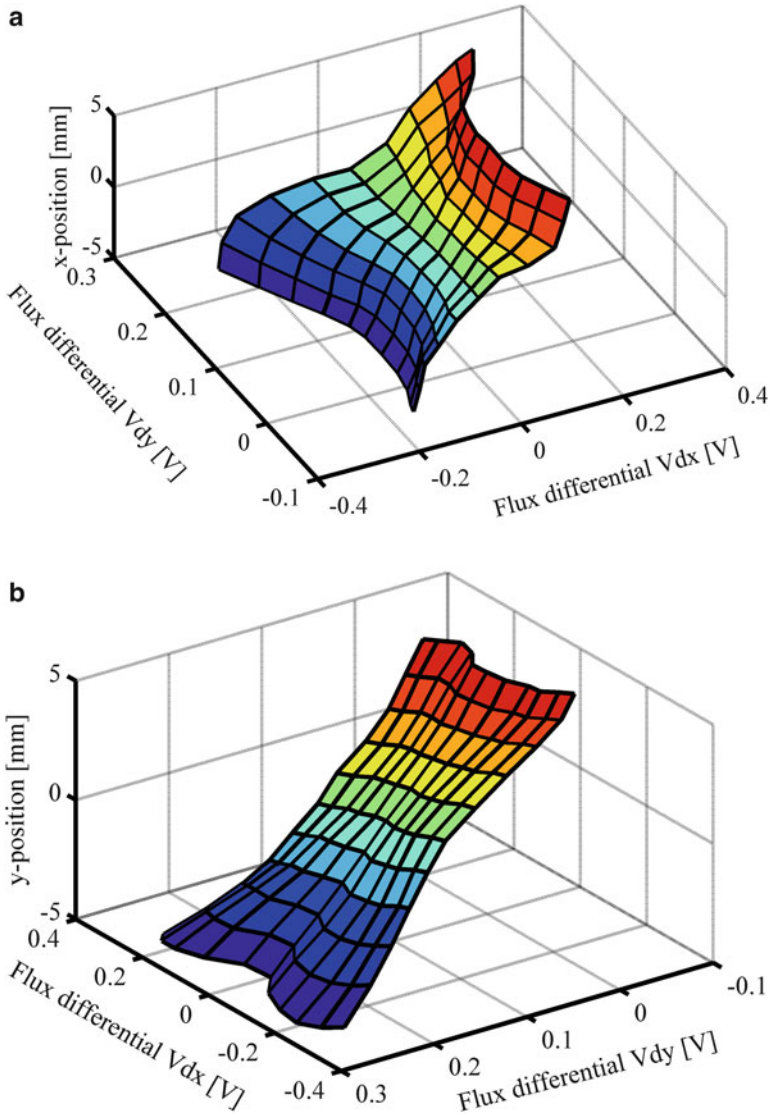
Fig. 27.11 Surface plot of measurements: (**a**) flux differential in x- and y- direction vs. x-position; (**b**) flux differential in x- and y-direction vs. y-position [38]

mth order polynomial of V_{dx} and the n_t order polynomial of V_{dy} were mapped into the x-direction position estimation function. In Eq. (27.17), the rth order polynomial of V_{dx} and the sth order polynomial of V_{dy} were mapped into the y-direction position estimation function.

$$
x = \begin{bmatrix} V_{dx}^m \\ \vdots \\ V_{dx}^0 \end{bmatrix}^{\mathsf{T}}_{1\times(m+1)}
\begin{bmatrix}
0 & \cdots & \cdots & 0 & p_{m0} \\
\vdots & \ddots & \ddots & p_{(m-1)1} & p_{(m-1)0} \\
\vdots & & \ddots & \ddots & \vdots & \vdots \\
0 & p_{1(n-1)} & \cdots & p_{11} & p_{10} \\
p_{0(n)} & p_{0(n-1)} & \cdots & p_{01} & p_{00}
\end{bmatrix}
\begin{bmatrix} V_{dy}^n \\ \vdots \\ V_{dy}^0 \end{bmatrix}^{\mathsf{T}}_{(n+1)\times 1}
\tag{27.16}
$$

$$
y = \begin{bmatrix} V_{dx}^r \\ \vdots \\ V_{dx}^0 \end{bmatrix}^{\mathsf{T}}_{1\times(r+1)}
\begin{bmatrix}
0 & \cdots & \cdots & 0 & q_{r0} \\
\vdots & \ddots & \ddots & q_{(r-1)1} & q_{(r-1)0} \\
\vdots & & \ddots & \ddots & \vdots & \vdots \\
0 & q_{1(s-1)} & \cdots & q_{11} & q_{10} \\
p_{0(s)} & p_{0(s-1)} & \cdots & q_{01} & q_{00}
\end{bmatrix}
\begin{bmatrix} V_{dy}^s \\ \vdots \\ V_{dy}^0 \end{bmatrix}^{\mathsf{T}}_{(s+1)\times 1}
\tag{27.17}
$$

where x and y are estimated positions using the magnetic field measurement. $V_{dx}^i (i = 1, 2, \ldots n(m))$ is the ith order of voltage differential in the x-direction. $V_{dy}^j (i = 1, 2, \ldots n(m))$ is the jth order of voltage differential in the y-direction. p_{ij}, q_{kl} are the coefficients of mapping functions. For example, p_{ij} is the coefficient of the product of ith order V_{dx} and jth order V_{dy} in the x-direction position determination function. Similarly, q_{kl} is the coefficient of the product of the kth order V_{dx} and lth order V_{dy} in the y-direction position determination function. The orders of the proposed polynomial models can be determined using cross-validation method by minimizing the root square mean error of fitting models.

3.2 Sensor Switching Control in Environments with Optical Blockage

The control system schematic with magnetic flux feedback for open-loop navigation of the microrobot on a horizontal plane is shown in Fig. 27.12.
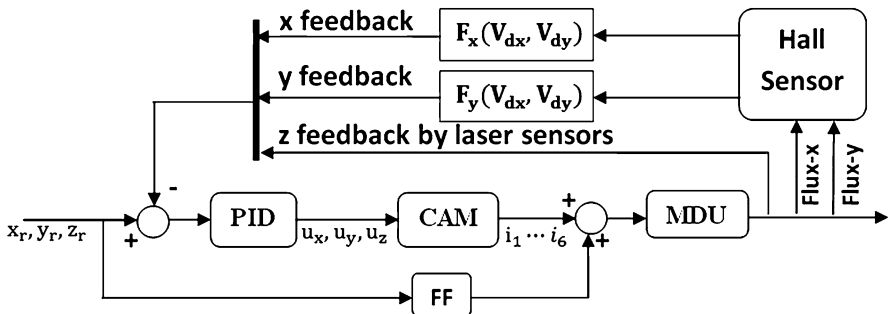
Fig. 27.12 Schematic diagram of proposed 2-D Hall-effect sensor feedback control [38]

Fig. 27.13 Configuration of laser sensors system. An object was placed in the workspace and blocked the laser beam of L2 [38]
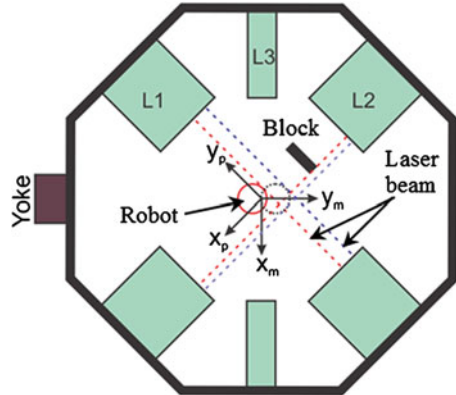
Mapping models were implemented in the control system to predict the *x*- and *y*-direction position of the microrobot. The Hall-effect sensors measured the magnetic field in the working space. The measurement was then fed to the polynomial mapping model. The mapping models calculated the *x*- and *y*-direction position based on the measured flux differential. Then the calculated position was fed back to the controller to form a closed loop position controller.

The proof-of-concept of switching the laser sensor based navigating mechanism to the magnetic flux based navigation method was evaluated experimentally. The configuration of laser sensors system of the maglev system is shown in Fig. 27.13. The laser sensors system is composed of three pairs of laser sensors L1, L2, and L3. L3 measures the *z*-direction position. L1 and L2 are used to measure the horizontal position of the microrobot.

Figure 27.14 shows the results of sensor switching experiment. In this experiment, the microrobot was navigated along *y* axis, and was kept at $x = 0$ mm in the *x* axis, at $z = 78$ mm in the Z_m axis. The position of the microrobot can be measured by the laser sensor at positive *y* axis. However, the magnetic flux position determination method switched on when the microrobot moved toward the negative *y* axis since the laser beam of L2 was blocked by the environment near the $y = 0$ mm location.

4 Off-Board Force Measurement and Its Application

4.1 Off-Board Force Measurement

In the steady state of a contact-free levitation, the levitated microrobot stays at the minimum magnetic potential energy point on a horizontal plane. This minimum magnetic potential energy point is the location of $B_{\max}$ Point $(x_{\max}, y_{\max})$. However, when the microrobot is in contact with its environment, the microrobot is not
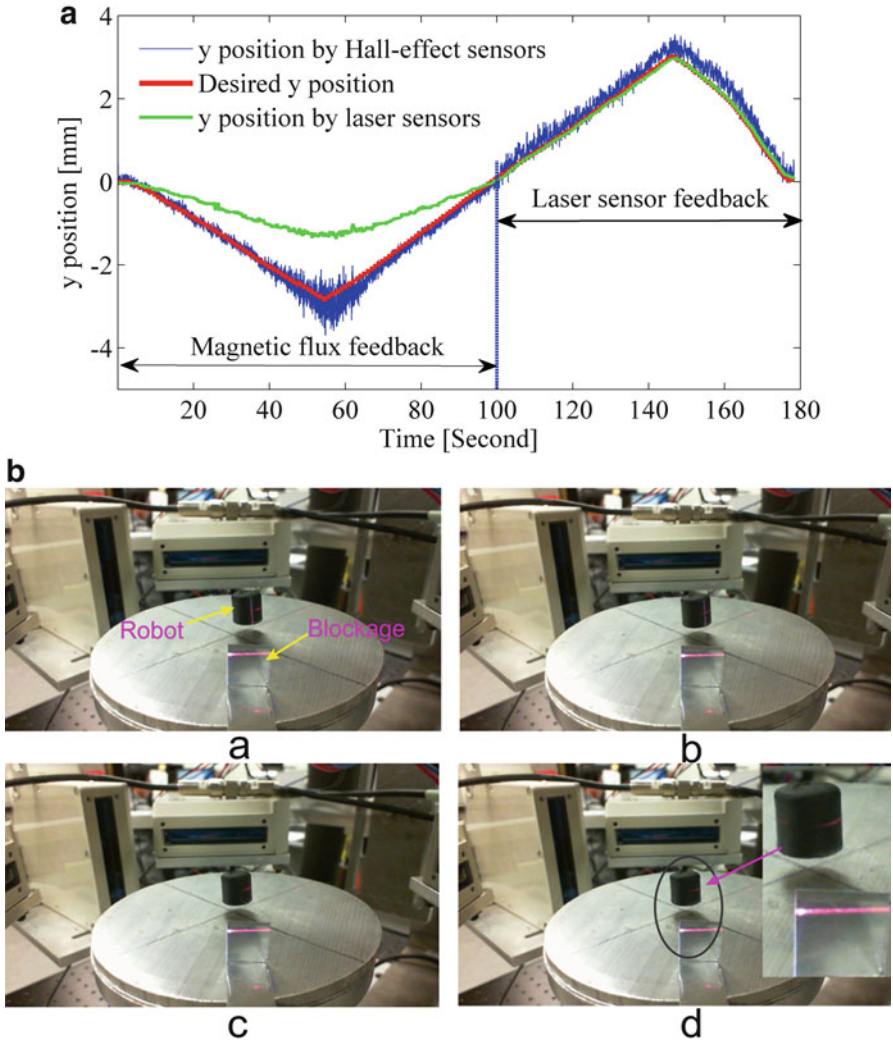
Fig. 27.14 Experiment result and (**a**) Sensor switching and motion tracking performance in y-direction; (**b**) Real time photo shot during the experiment

stabilized at the B_{max} position. The environment exerts contact force on the microrobot. Meanwhile, the magnetic field produced by the levitation stage exerts an equal and opposite magnetic force that moves the microrobot toward the B_{max} point. Figure 27.15 explains this fundamental. Due to the tiny size of the levitated microrobot and the remote manipulation property of levitation, it is very difficult to install an on-board force to measure the contact force on the microrobot. In addition, it is not energy efficient to levitate extra mass.
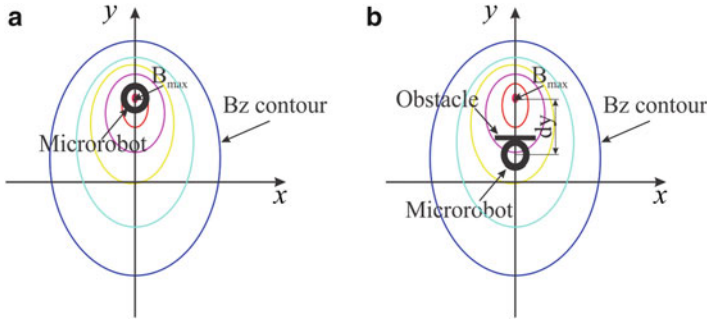
Fig. 27.15 Location of levitation robot and position of B_{max} point. (**a**) coincide; (**b**) separated

A novel off-board force determination mechanism for magnetic navigation was proposed to determine the real contact force and assist maintaining a stable levitation. For the proposed method, no force sensor is attached on the microrobot, and the cost is extremely low. The force determination mechanism was validated experimentally in a single direction in [39, 40]. The principle is that the contact force on the microrobot is linear to the distance between the real position of the levitated microrobot and the position of the B_{max} point. The force model in y is an example, where the force model is expressed as

$$F_y = C(y_r - y_{max}) \tag{27.18}$$

where y_r is the real position of the microrobot in the y-direction, $y - max$ is the position of the B_{max} point in the y-direction, and C is a constant similar to the stiffness of a mechanical spring. The B_{max} position, as presented in the previous section, is determined using Hall-effect sensors attached at the bottom of the pole-piece. The real position of the microrobot is measured using laser-beam sensors. It was reported in [39] that the force determination method provides $1.27\,\mu N$ accuracy for single axis force sensing.

Experimental result is presented in Figs. 27.16 and 27.17 to show the performance of the proposed off-board force determination mechanism. In the experiment, the microrobot was navigated to push an aluminum cantilever beam step-by-step in the y-direction. The real contact force was measured by measuring the deflection of the cantilever beam using a high accuracy laser sensor. It shows that the magnetic flux measurement based off-board force mechanism can provide very accurate force measurement. It should be noticed that the output force of a magnetically navigated microrobot is determined by the volume of magnetic material on the microrobot. A larger microrobot provides more pushing force on its environment.
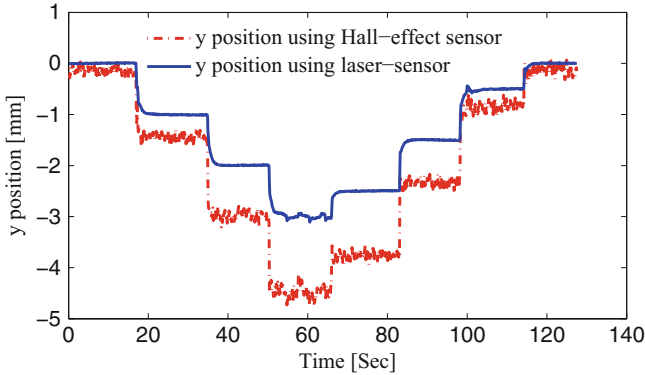
Fig. 27.16 Trajectory of B_{max} measured by Hall-effect sensors and trajectory of the microrobot measured by laser beam sensors
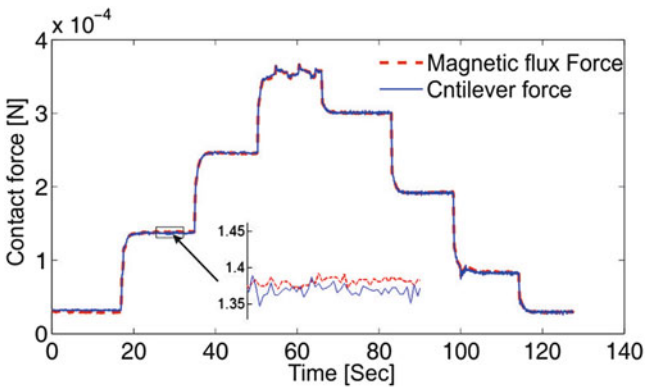
Fig. 27.17 Compare between measured force based on cantilever deflection and calculated force based on distance between B_{max} location and microrobot real position

4.2 Remote Manipulation for Biomedical Applications

Biomedical applications, such as medical surgery and drug delivery, have been important contributions of MNM. Traditional surgery processes involve cutting special parts of patient or using soft pipe in order to access the organisms that require an operation. These procedures make patients feel uncomfortable and usually require a long recovery period. In addition, taking drugs through mouth reduces the effectiveness of medicine. MNM, with the feature of wireless control and the potential to be down scaled, have significant advantages over traditional manipulators. For instance, a microrobot can be sent directly to the organism that need a surgery through the mouth or vessel, which prevents damaging other parts of human body and is less uncomfortable [41]. Also, delivering medicine directly to the target position can improve the efficiency of drugs.
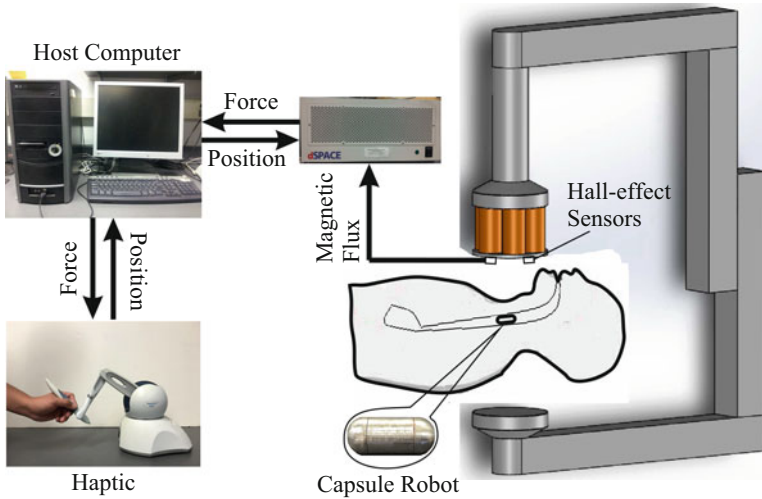
Fig. 27.18 Potential biomedical application: remotely navigate a capsule robot into human body with force sensing at the master side

Most MNM for biomedical applications are passively actuated without active position or force control. This is owing to the difficulty of localizing the microrobot in an out-of-reach environment. Moreover, the tiny size of the microrobot makes it impossible to attach an on-board force sensor to measure the applied force on the microrobot by its environments, thus reducing the success rate of manipulation.

With the off-board force sensing method presented in the last section and for the proof-of-concept, a remote micromanipulation system that combines the developed magnetic levitation stage and a Haptic device was proposed for potential applications such as drug delivery and explorations of the gastrointestinal (GI) using a video capsule endoscopy [40]. The system is shown in Fig. 27.18.

The system has a master side and a slave side. The Haptic device, the master side, generates the motion trajectory of the microrobot. The magnetically navigated capsule robot [42, 43], the slave side, follows the commanded motion trajectory. When the microrobot is in contact with its environment, the system improves the stability of magnetic levitation by transferring the micro force measured at the microrobot side using Hall-effect sensors to the macro force on the operator side.

5   Conclusion

This chapter presents a magnetically driven micromanipulation system. The principle of navigation, modeling the magnetic field, and navigation control of a microrobot were presented. A PID controller was proposed to control the motion of the microrobot with $20\,\mu\text{m}$ accuracy. However, cross-coupling effect occurred

during the navigation. One challenge of magnetic navigation is localizing the micro-robot. A sensor switching motion control mechanism was proposed to maintain the persistent navigation of a microrobot when optical blockage exists. The system proposed in this chapter has potential applications in biomedical surgery and drug delivery. An off-board force sensing mechanism using magnetic flux measurement was developed to stabilize the navigation.

References

1. Ji L, Xu L, Jin C (2013) Research on a low power consumption six-pole heteropolar hybrid magnetic bearing. IEEE Trans Magnet 49(8):4918–4926
2. Chen S, Hsu C (2002) Optimal design of a three-pole active magnetic bearing. IEEE Trans Magnet 38(5):3458–3466
3. Perez-Diaz et al (2014) Superconducting noncontact device for precision positioning in cryogenic environments. IEEE/ASME Trans Mechatron 19(2):598–605
4. Schultz L et al (2005) Superconductively levitated transport system: the supratrans project. IEEE Trans Appl Supercon 15(2):2301–2305
5. Simon MD, Geim AK (2000) Diamagnetic levitation: flying frogs and coating magnets. J Appl Phys 87(9):6200–6204
6. Lyuksyutov I, Naugle G, Rathnayaka K (2004) On-chip manipulation of levitated femto-droplets. Appl Phys Lett 85(10):1817–1819
7. Guo Y et al (2007) Design and analysis of a prototype linear motor driving system for HTS maglev transportation. IEEE Trans Appl Supercon 17(2):2087–2090
8. Trumper D, Kim W, Williams M (1996) Design and analysis framework for linear permanent-magnet machines. IEEE Trans Ind Appl 32(2):371–379
9. Covert E, Finston M, Vlajinac M, Stephens T (1973) Magnetic balance and suspension systems for use with wind tunnels. Prog Aerosp Sci 14:27–94
10. Yu W, Li X (2014) A magnetic levitation system for advanced control education. IFAC Proceedings Volumes, 47(3):9032–9037
11. Souza G et al (2010) Three dimensional tissue culture based on magnetic cell levitation. Nat Nanotechnol 5:291–296
12. Liu J et al (2015) Automated vitrification of embryos: A robotics approach. IEEE Robot Autom Mag 22(2):33–40
13. Karimirad F, Shirinzadeh B, Yan W, Fatikow S (2013) A vision-based methodology to dynamically track and describe cell deformation during cell micromanipulation. Int J Optomechatronics 7:33–45
14. Avci E et al (2013) Toward high-speed automated micromanipulation. In: IEEE international conference on robotics and automation
15. Wang D, Yang Q, Dong H (2013) A monolithic compliant piezoelectric-driven microgripper: design, modeling, and testing. IEEE Trans Mechatron 18(1):138–147
16. Xun X et al (2015) A novel piezo-driven microgripper with large jaw displacement. Microsyst Technol 21:931–942
17. Alleyne A et al (2013) Motion control for magnetic micro-scale manipulation. In: European control conference
18. Mehrtash M, Khamesee MB (2013) Modeling and analysis of eddy-current damping effect in horizontal motions for a high-precision magnetic navigation platform. IEEE Trans Magnet 49(8):4801–4810
19. Kim W, Trumper D, Lang J (1997) Modeling and vector control of a planar magnetic levitator. In: IEEE industry applications conference, pp 349–356

20. Chen M, Lin T, Hung S, Fu L (2012) Design and experiment of a macro-micro planar maglev position system. IEEE Trans Ind Electron 59(11):4128–4139
21. Busch-Vishniac IJ (1990) Applications of magnetic levitation-based micro-automation in semiconductor manufacturing. IEEE Trans Semicond Manuf 3(3):109–115
22. Mehrtash M, Tsuda N, Khamesee MB (2011) Bilateral macro-micro teleoperation using magnetic levitation. IEEE/ASME Trans Mechatron 16(3):459–469
23. Peyer K, Zhang L, Nelson BJ (2013) Bio-inspired magnetic swimming microrobots for biomedical applications. Nanoscale 5:1259–1272
24. Martel S et al (2009) Flagellated magnetotactic bacteria as controlled MRI-trackable propulsion and steering systems for medical nanorobots operating in the human microvasculature. Int J Robot Res 28(4):571–582
25. Grady MS et al (1990) Nonlinear magnetic stereotaxis: three-dimensional in vivo remote magnetic manipulation of a small object in canine brain. Med Phys 17(3):405–415
26. Kummer MP et al (2010) OctoMag: an electromagnetic system for 5-DOF wireless micromanipulation. IEEE Trans Robot 26(6):1006–1017
27. Carpi F, Pappone C (2009) Magnetic maneuvering of endoscopic capsules by means of a robotic navigation system. IEEE Trans Biomed Eng 56(5):1482–1490
28. Weizenecker J et al (2009) Three-dimensional real-time in vivo magnetic particle imaging. Phys Med Biol 54(5):1–10
29. Steager EB et al (2013) Automated biomanipulation of single cells using magnetic microrobots. Int J Robot Res 32(3):346–359
30. Gosse C, Croquette V (2002) Magnetic tweezers: micromanipulation and force measurement at the molecular level. Biophys J 82:3314–3329
31. Nakamura T, Khamesee MB (1997) A prototype mechanism for three-dimensional levitated movement of a small magnet. IEEE/ASME Trans Mechatron 2(1):41–50
32. Shameli E, Craig DG, Khamesee MB (2006) Design and implementation of a magnetically suspended microrobotic pick-and-place system. J Appl Phys 99:08P509
33. Khamesee MB, Kato N, Nomura Y, Nakamura T (2002) Design and control of a microrobotic system using magnetic levitation. IEEE/ASME Trans Mechatron 7(1):1–14
34. Khamesee MB, Shameli E (2005) Regulation technique for a large gap magnetic field for 3D non-contact manipulation. Mechatronics 15(9):1073–1087
35. Mehrtash M, Khamesee MB (2011) Design and implementation of LQG/LTR controller for a magnetic telemanipulation system- performance evaluation and energy saving. Microsyst Technol 17(5):1135–1143
36. Islam SM (2014) Magnetic-based closed-loop control of paramagnetic microparticles using ultrasound feedback. In: IEEE international conference on robotics and automation (ICRA), pp 3807–3812
37. Mehrtash M, Khamesee MB Tsuda N, Chang J (2011) Motion Control of a magnetically levitated microrobot using magnetic flux measurement. Microsyst Technol 18(9):1417–1424
38. Zhang X, Mehrtash M, Khamesee MB (2015) Dual-axial motion control of a magnetic levitation system using hall-effect sensors. IEEE/ASME Trans Mechatron. doi:10.1109/TMECH.2015.2479404
39. Mehrtash M, Khamesee MB (2013) Micro-domain force estimation using hall-effect sensor for a magnetic microrobotic station. J Adv Mech Des Syst Manuf 7(1):2–14
40. Mehrtash M, Zhang X, Khamesee MB (2015) Bilateral magnetic micromanipulation using off-board force sensor. IEEE/ASME Trans Mechatron 20(6):3223–3231
41. The Economist Magazine (April-May 2009), Science and technology section: look, no wires—a tiny, levitating robot takes to the air, p 85
42. Hosseini S, Mehrtash M, Khamesee MB (2011) Design, fabrication and control of a magnetic capsule robot for the human esophagus. J Microsyst Technol 17(5):1145–1152
43. Gumprecht JD, Lueth TC, Khamesee MB (2013) Navigation of a robotic capsule endoscope with a novel ultrasound tracking system. Microsyst Technol 19:1415–1423

Chapter 28
Design, Fabrication, and Robust Control of Miniaturized Optical Image Stabilizers

Kaiwen Yuan, Alireza Alizadegan, Pan Zhao, Ryozo Nagamune, Jingsong Chu, Simon Park, and Mu Chiao

Abstract 3D printing and moulding (3DPM) method is applied to the fabrication of a miniature magnetic actuator for optical image stabilization (OIS) applications. Polydimethylsiloxane (PDMS) and strontium ferrite (SrFe) nano powder are used as the main structural materials. Young's modulus and the magnetization of the material with SrFe-doping ratios ranging from 20 to 60 % by weight are characterized. The actuator, consisting of four coils, an actuating plate, and a base supporter, is assembled and tested with a laser Doppler velocimetry (LDV) system. A tilting angle of 0.6° is achieved with the application of 500 mA (50 turns/9 mm long coils). A Taguchi's orthogonal experimental design is used in the finite element analysis (FEA) simulation to examine the effect of dimension variations on the eigenfrequencies. The frequency response of the actuator is characterized, and the experimental results match with the simulation results between 1 and 450 Hz showing less than 5 % errors and bandwidth of around 56 Hz. A series of replica experiments are also performed and analyzed. Additionally, unavoidable product variabilities in the batch fabrication of miniature actuators are discussed, and the application of robust control to deal with such variations is proposed.

Keywords Optical image stabilizer • 3D printing and moulding • Miniature magnetic actuator • Polydimethylsiloxane • Strontium ferrite • Microelectromechanical systems • Finite element analysis • Batch fabrication • Product variability • Frequency response • Replica experiment • Robust control

K. Yuan • A. Alizadegan • P. Zhao • R. Nagamune (✉) • M. Chiao
Department of Mechanical Engineering, University of British Columbia,
Vancouver, BC, Canada V6T 1Z4
e-mail: muchiao@mech.ubc.ca; nagamune@mech.ubc.ca

J. Chu
Micromolding Solutions Inc., 9595 Rue Ignace, Suite 200, Brossard, QC, Canada J4Y 2P3
e-mail: jingsong.chu@micromolding.ca

S. Park
Schulich School of Engineering, University of Calgary, 2500 University Drive NW,
Calgary, AB, Canada T2N 1N4
e-mail: sipark@ucalgary.ca

1 Introduction

3-Dimension (3D) printing has been applied to rapid-prototyping and customization with advantages of low cost, speed, and ease of fabrication [4] in many fields, including medical applications and consumer electronics. 3D printing relies on an additive manufacturing process to build products layer-by-layer with software that processes computer-aided design (CAD) files. Sharing the same mechanism, 3D printers can print in multiple ways, such as with selective laser sintering (SLS), fused deposition modeling (FDM), stereolithography (SLA), and so on [3]. Diverse printing materials like acrylonitrile butadiene styrene (ABS), polylactic acid (PLA), polycarbonate (PC), polyphenylsulfone (PPSU), and nylon can be used. High-resolution 3D printing methods such as stereolithography (SLA) can allow 3D printing to rapidly fabricate devices and parts with a high precision [4]. Nevertheless, in comparison with traditional injection molding and other technologies, 3D printing is not economical for large-scale production [6] and the materials for high resolution 3D printing are less diverse. For example, an SLA printer can only use photopolymer resin that is solidified with an ultraviolet laser.

The application of 3D printing in injection moulding tooling (i.e., moulds and cooling channels) fabrication [40] and in vacuum-assisted resin transfer moulding [14] already shows advantages in both the quality and rate. In traditional moulding methods, for fabricating a mould, generally expensive, computer numerical control (CNC) end mills are required, which are both costly and brittle, especially for micro- or miniature-scale cases. For micro-mould fabrication, UV photolithography with clean-room conditions is always required [45], which limits the production efficiency and adds to the cost. The combination of moulding and 3D printing (for fabricating the toolings for the moulding methods) provides a scalable, faster alternative at a lower cost than either traditional moulding methods or micro-mould fabrication.

Several attempts have been made to use high-resolution 3D printing moulding (3DPM) in lab-on-chip (LOC) [11, 23] and microlens fabrication [42]. The efforts simplify the micro-fabrication process, shorten the machining duration, and reduce the cost. Nevertheless, so far, little research has been conducted on 3DPM applications for actuator fabrication or magnetic material, which require more considerations such as actuator assembly and accessory components to magnetize the material.

Generally, an actuator is one of five components for any control system, which also includes plant, sensor and transducer, signal modification units, and controllers [13]. Magnetic actuators are an important type of actuators, and they have been applied to drug delivery [1, 34], disk drives [7], cellular migration [26], and so on. Compared to electrostatic actuators, magnetic actuators have advantages in operating wirelessly and at lower voltage.

To achieve and control both desired mechanical and magnetic properties, magnetic particles should be loaded into polymers to form composites and be actuated

with coils. By adding strontium ferrite micropowder into a commercial epoxy resin and actuating it with a planar coil, large deflections are possible [28, 29]. Pallapa et al. [36] used lithography and micromoulding techniques to mix Nd-Fe-B powder into a PDMS matrix and the actuator was actuated with printed circuit board planar coils. In other experiments, a micromirror scanner was demonstrated using a maximum 20 mT magnetic field [36]. In any case, few authors have reported adding strontium ferrite nanopowder into a PDMS to form actuators. PDMS, a porous, biocompatible (non-toxic) and easy to process material is preferable for the base material and convenient for post-processing (i.e., laser cutting and coating). Strontium ferrite powders also have advantages over other hard magnetic material candidates of low cost and chemical inertness [18]. Furthermore, for most miniature or micro-magnetic actuators, fabrication processes require clean-room conditions [16, 26, 36, 37], which is both time-consuming and costly. Therefore, innovative methods need to be developed that simplify the fabrication process and reduce the cost.

Image stabilization (IS) is a good solution to correct the image blur that results from a photographer's hand trembling. Optical image stabilization (OIS) and digital image stabilization (DIS or EIS) are two types of popular IS methods. DIS shifts the electronic image from frame to frame to counteract the motion [8] or uses filters (i.e., Kalman filters) to track the movement of pixels and remove detected irregular global motion effects [15]. Although the cost of DIS is low, since no physical components like a gyroscope or actuators are added, it is less effective and may sacrifice resolution. Therefore, OIS, which modifies the optical path by an actuator to compensate for hand-tremble, is more advantageous and promising [30, 44]. A typical OIS system includes a gyroscope to measure hand tremor, an image sensor, an actuator to modify the optical path, and a controller to process signals and give commands to the actuator. Lens shifting, module tilting, and image sensor shifting are commonly used technologies for OIS systems [39, 44]. Actuating methods have been published, such as the voice coil motor (VCM) [10, 27, 44, 46], moving mirrors methods [20], piezoelectric (PZT) methods [25], and microelectromechanical system (MEMS) based on thermal actuating methods [9, 31]. VCM applies coils to actuate magnets, which are usually bonded with the lens holder, so that actuating displacement can be controlled by the current/voltage in the coils. As VCM has advantages in its low operating voltage, cost effectiveness, and stability, it has been the main method applied in the OIS industry [21, 43]. Our previous publication [38] reported an innovative tilting lens VCM actuator concept and we demonstrated a conception macro prototype. As a follow-up to that research, we fabricated and tested a miniature prototype.

For OIS applications, the desired performance of the actuating system should include the actuator tilting in yaw (or pitch) direction of 0.6° [17] with a low energy consumption. We applied high-resolution 3D printing moulding (3DPM) technology to fabricate a magnetic actuator, according to the above specifications and researched the properties of a mixture of strontium ferrite nanopowder and PDMS as the filling material, to help us understand and predict the dynamic and stationary performance of the actuator. Influences of dimension variations

on performance were studied, especially dynamic performance since errors in fabrication can lead to variations in the dimensions of parts. These can affect the performance.

The fabricated actuators have product variations due to the tolerance in fabrication process. Such product variations can be seen in the resonant frequencies and DC gains in frequency responses. To achieve the actuator's fast and accurate performance uniformly over all actuators, we propose to utilize the robust control techniques based on a mathematical model with parameter uncertainties [24, 35, 47].

2 Design and Fabrication of Lens-Tilting OISs

In this section, we present the design of the actuator in lens-tilting OISs as well as the fabrication details and the replica experiments.

2.1 Design

A final actuating system (Fig. 28.1) consists of four parts: (1) actuating plate (or lens holder, with a ring magnet attached), (2) lens, (3) actuating coils (with Ni-Zn cores inside), and (4) a base supporter. In this paper, we focus on parts 1, 3, and 4. The design refers to previous work from our group [32, 38] that used coils to actuate a platform (the lens holder). Since folded beams [32] may cause difficulties in the demoulding, the design with one beam on each side was used to directly connect the parallel beams on the supporter and the center circle.

Figure 28.2 shows the dimensions of the actuating plate; the thickness is 1 mm. The entire length of the system is 12 mm, and four pairs of beams are folded at an

Fig. 28.1 Exploded view of the actuator system, composed of lens holder, coils, and supporter
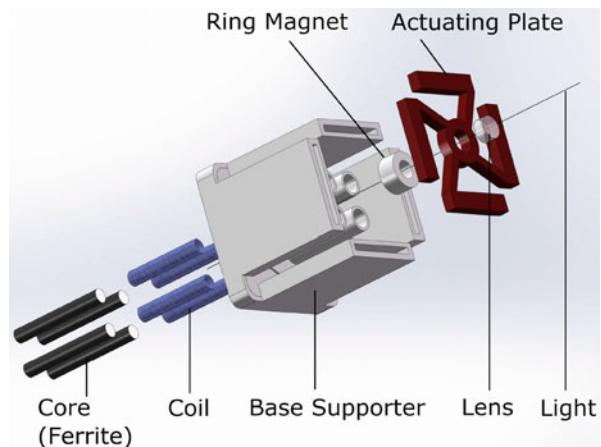
Fig. 28.2 *Top view* and dimensions of the actuating plate
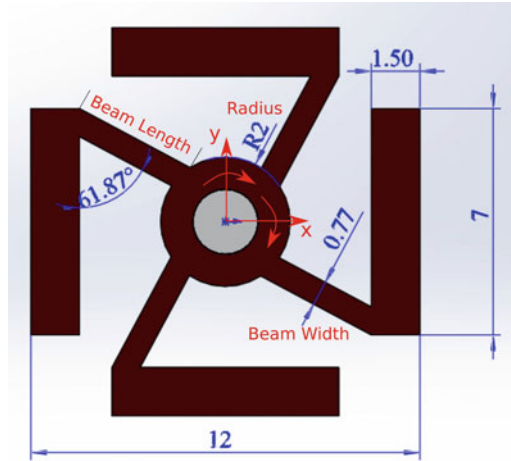
angle of 61.87° and a 2 mm radius central circle. All coils are wound with wires in the same direction. The coil has 50 turns and has a Ni-Zn core inside, with a length of 9 mm and an outer diameter of 1.35 mm.

When applying current with the same direction as the four coils, the force between the coils and the lens holder (including ring magnet and actuating plate) will cause the plate to translate vertically. When an opposite current is applied to two pairs of coils on the same side, the plate will rotate in yaw (around the *x* axis; Fig. 28.2) or pitch (around the *y* axis; Fig. 28.2). If a hole is drilled in the center of the plate, the lens can be mounted and when the plate rotates the light path will change. By changing the light path, the actuator can compensate for the hand-trembling effect with the help of sensors and controller. For hand tremor, the frequency can be as low as about 8–12 Hz [17].

2.2 Fabrication

In the following, details including material preparation, fabrication process flow (both 3D printing and moulding), and replica experiments are described.

2.2.1 Material Preparation

A layer of 20 % weight/volume polyacrylic acid (PAA, Sigma-Aldrich Corporation) solvent (in water) was spin-coated onto glass slides as a sacrificial layer at 2000 rpm for 30 s and then the PAA-coated glass slides were baked at 120 °C for 3 min. PDMS (Sylgard 184 Silicone Elastomer, Dow Corning Corporation) was used as the base material. Pre-polymer and crosslinker, two compounds of PDMS, were mixed at

Fig. 28.3 Mould of actuating
plate for stereolithography 3D
printing
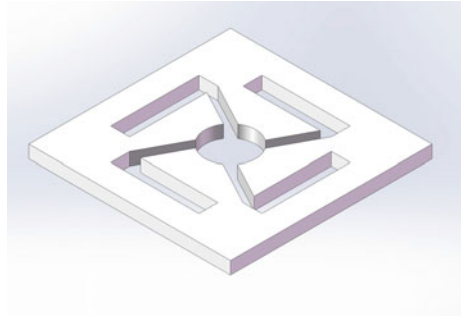
a 5:1 ratio by weight. Subsequently, strontium ferrite nanopowder (Sigma-Aldrich
Corporation) was introduced at 3:2 (60 %), 2:3 (40 %), or 4:1 (20 %) (strontium
ferrite:PDMS) by weight, and mixed with a stick for at least 1 min.

Stereolithography 3D printing (SLA 3DP) [22] is a 3D printing method that
can achieve high resolution and it was applied in our project. The mechanism of
SLA 3DP uses UV light to solidify the photopolymer, layer by layer, so that the
parameters such as slice thickness, burn-in exposure time, and normal exposure time
are significant.

Asiga Freeform Pico (ASIGA, CA, USA), an affordable stereolithography (SLA)
3D printer with native XY pixel resolutions down to 27 μm and 250 nm Z-axis servo
resolution, was used to print the mould, with PlasWHITE (photopolymer, ASIGA,
CA, USA) as the material. A flash mould (without bottom; Fig. 28.3) was designed
to ease the demoulding process, which typically has the most difficulties [19] due
to large surface energy and friction force. After printing, the printed mould was
immersed into isopropyl alcohol (IPA) for 5 min to remove uncured photopolymer.
When dry, it was cured for 20 min in UV light (Pico Flash).

2.2.2 Moulding

The fabrication procedures (Fig. 28.4) for the actuating plate are as follows: (1) 3D
printed mould mounted on PAA-coated side of the glass slide; (2) cylinder magnet
(MAG400190, Main Electronic Supplies Ltd.) positioned at the bottom of the slide;
(3) mixture of PDMS and strontium ferrite filled into the mould; (4) mould carefully
covered with another PAA-coated glass slide and then clamped (Fig. 28.5) with
another cylinder magnet (MAG400185, Main Electronic Supplies Ltd), which also
helps the magnetic powder align vertically during the curing process; (5) curing the
moulded sandwich at 75 °C for 3 h (on a hot plate or in the oven) and then natural
cooling; (6) magnets slowly removed and the cooled mould sandwich is immersed
in water for 5 min to dissolve the PAA sacrificial layer; (7) slides are uncovered to
obtain the mould with the solidified part inside; and (8) demoulding with tweezers
or a thin stick.

Fig. 28.4 Fabrication process for the actuating plate

Fig. 28.5 Moulding
sandwich with magnets
clamping glass slides and
magnetizing materials in
mould
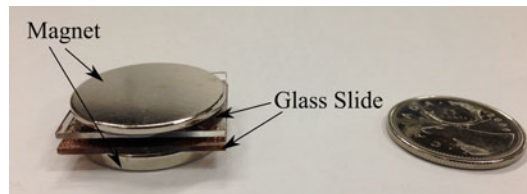
The 20 and 40 % SrFe-doped mixtures (Fig. 28.6b, c) have viscosities that are
low enough to stay in the liquid phase. In step (3), we just need to drop the liquid
mixtures into the channels or the center and the liquid will automatically flow to fill
the mould. Nevertheless, for the 60 % SrFe-doped mixture (Fig. 28.6d), the mixture
is mud-like (higher viscosity) and to fill the mould, flat tools, such as a glass slide,
must be used to help fill and compress the materials into the mould. Although it
seems to be more complicated, the 60 % mixture is actually more advantageous than
the 20 or 40 % mixtures for the mould (or flash mould) we used, because the liquid
mixtures tend to flow out, resulting in a waste of material and a failure to completely
fill the mould. By attaching a magnet to the back of the glass slides during filling
(Figs. 28.4–3), the filling material will be attracted to the bottom because of its
magnetic properties, and the magnet will cause a virtual external compression, to
increase the density and amount of SrFe in the mould volume.

After obtaining the actuating plate, the base supporter was also printed using the
SLA 3D printer. The post-processing was the same as that of the moulding (i.e.,
printed part immersed in IPA for 5 min to remove uncured photopolymer, and after
drying, the curing takes place for 20 min with UV light (Pico Flash)). To assemble
the device (Fig. 28.7, the coils (Senders Electronic Commerce Co., Ltd., with HP06
core) were first inserted into the four hollow cylinders with wires coming out. Then,
a ring magnet (R211, K&J Magnetics Inc.) was attached coaxially to the center of
the actuating plate. When placing the plate onto the glass slide, a drop of PDMS
(pre-polymer and crosslinker mixed with 5:1 ratio by weight) was dripped into the
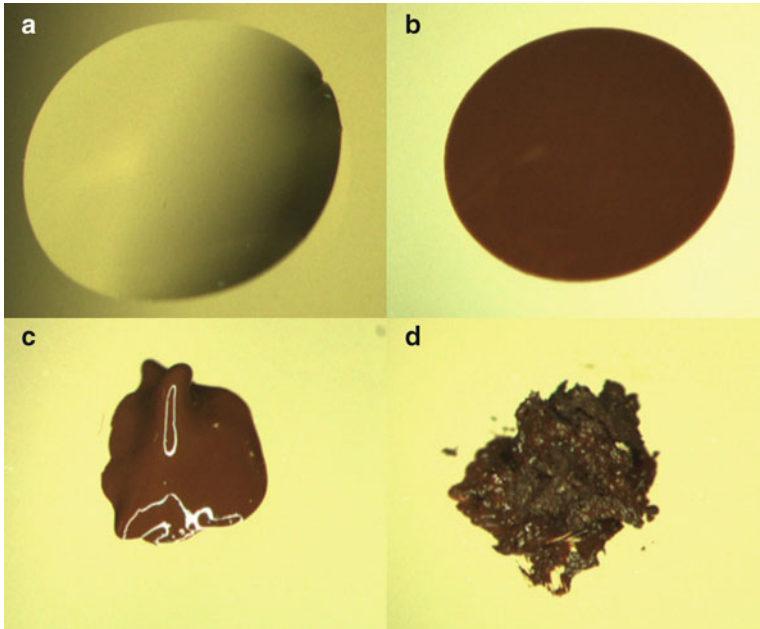
Fig. 28.6 (**a**) A drop of PDMS (pre-polymer and crosslinker mixed with 5:1 w/w) on glass slide; (**b**) a drop of PDMS doped with 20 % w/w SrFe nanopowder on glass slide; (**c**) a drop of PDMS doped with 40 % w/w SrFe nanopowder on glass slide; (**d**) a drop of PDMS doped with 60 % w/w SrFe nanopowder on glass slide

hole of the ring magnet. Afterwards, the plate was cured at 75 °C for 40 min to solidify the PDMS, and to further bond the ring magnet and the plate together. If the plate is directly mounted on the supporter, the strong attractive force between the core of the coils and the ring magnet will cause the plate to snap immediately. Therefore, we filled PDMS (5:1) into the four grooves at the top of the walls of the supporter as a glue to bond the plate and supporter. Before mounting the plate, and to protect it from snapping, another R311 ring magnet was attached on the other side of the glass slide (Fig. 28.7, step 4; and Fig. 28.8a). Finally, the plate is mounted on top of the supporter, which is soldered with a flexible printed circuit board (FPCB) for testing. The device was cured at 75 °C for 1 h, and afterwards, the auxiliary parts were removed carefully. Figure 28.8b shows the final actuator after curing. A hole in the center of the plate can be cut with a laser easily or a cylinder can be inserted during moulding to embed any lens. Because PDMS is also a promising optical material [42], in the future, we may be able to mould both the lens and the lens holder directly. In this project, for ease of characterization, we kept the center material to emulate the case with a lens.
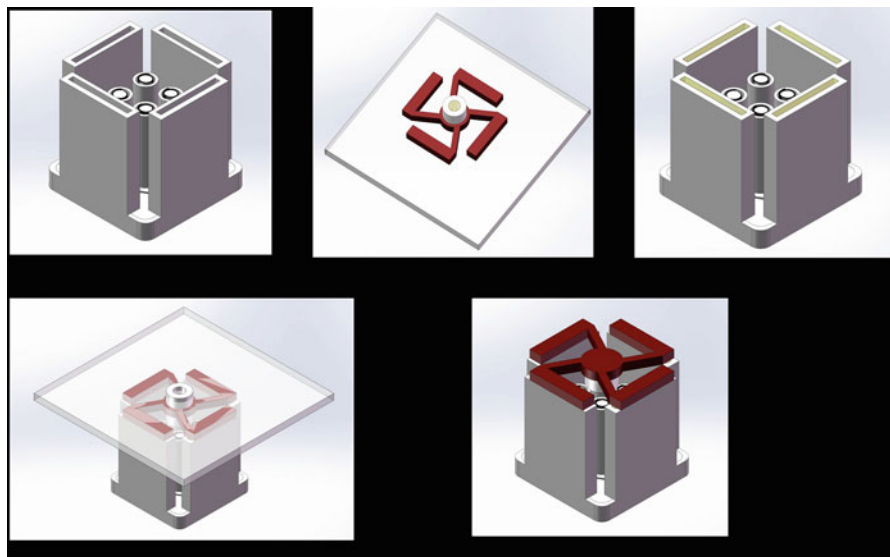
Fig. 28.7 Assembling process of actuator
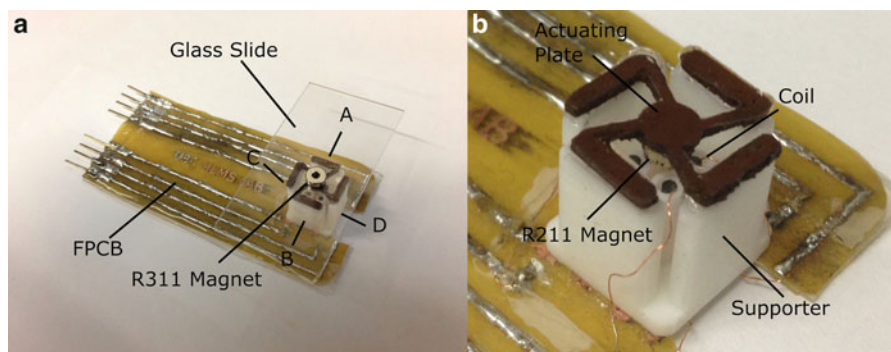
Fig. 28.8 (**a**) Actuator with auxiliary magnet and glass slide before curing; (**b**) final actuator after curing

2.2.3 Replica Experiments

A series of replica experiments were performed to study the influences of fabrication errors on variations of frequency responses of actuators. Since air gap, which may have significant influences on magnitude/tilting angle [38], but negligible impacts on shape of frequency response curves, is not easy to be quantified in assembled actuators, in this part, we mainly focus on studies of bandwidths and locations of natural frequencies. Firstly, five moulds (named as Mould1, Mould2, Mould3, Mould4, and Mould5) were printed with the same SLA 3D Printer. Each mould was
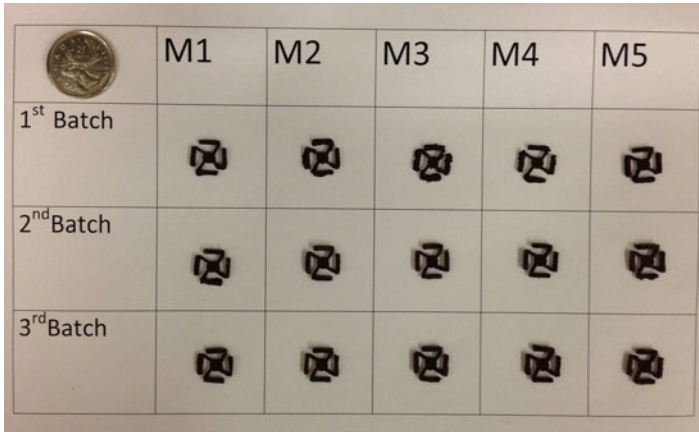
Fig. 28.9 Fifteen actuating plates with five moulds and three batches
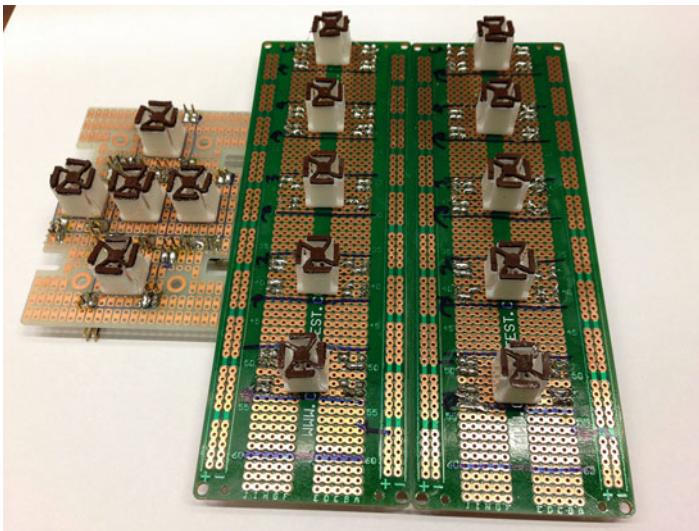
Fig. 28.10 Fifteen assembled actuators with five moulds and three batches (*from left to right*: first batch, second batch, and third batch)

used to replicate three actuating plates. The exactly same procedures were followed to fabricate 15 actuating plates in total, as shown in Fig. 28.9. Fifteen actuators based on these plates were assembled, as shown in Fig. 28.10. A frequency response test was conducted for the 15 assembled actuators, with sinusoidal excitations of fixed amplitude 100 mA and frequency range from 1 to 100 Hz.

Table 28.1 Dimension of samples for SQUID test

Ratio of Strontium Ferrite	Length (mm)	Width (mm)	Thickness (mm)	Mass (g)
20 % weight/weight	2.57	2.2	1	0.00681
40 % weight/weight	2.09	2.35	1	0.00692
60 % weight/weight	2.29	2.41	1	0.01061

3   Characterization Methods

To understand the material properties and obtain parameters for further simulation, a tensile test of material and magnetization test were performed to calculate Young's modulus, coercivity, remanent magnetic moment (unit weight), etc. For actuator performances, both stationary and dynamic tests were done with a laser Doppler vibrometer (LDV) system.

3.1   Characterization of Material

Sample preparation procedures were similar (but not identical) to those above: (1) PDMS (5:1 ratio of pre-polymer and crosslinker by weight) was mixed with strontium ferrite with 20, 40, and 60 % strontium ferrite:PDMS ratios by weight; (2) mixture was stirred with a stick for at least 1 min in petri-dishes (since the viscosity of the 60 % mixture is low, even using a glass slide to help smear the mixture, obtaining a uniform layer was difficult); (3) the sample was placed into an oven at 75 °C for 3 h with MAG400350 (Main Electronic Supplies Ltd) at the bottom, and followed with natural cooling; (4) samples were cut with a knife for characterizations. A Wyko profiler (VEECO Metrology Group, AZ, USA) was used to measure the thickness of samples and digital calipers were used to measure their length and width. The tensile test of composites were measured with a thermo-mechanical analyzer (TMA 2940-Q series, TA Instruments, DE, USA). Testing was performed at room temperature and the applied force was ramped with 0.05 N min^{-1} from 0 to 0.1 N or 0.2 N. After obtaining the data for each sample, the end of pre-loading was set as the start point. Strains and stresses were obtained by dividing the dimension changes by length (including pre-loading length) and dividing the forces by section areas. Young's modulus was calculated with the slope of each curve; the effect of temperature on Young's modulus was ignored here.

A superconducting quantum interference device (SQUID) (Quantum Design, CA, USA) was used to characterize the magnetic moment (emu) versus the applied magnetic field (kOe) at 300 °K (26.86 °C). Samples (Table 28.1) were prepared exactly as described for the moulds in the Fabrication section.

3.2 Characterization of Actuator

A LDV system (Fig. 28.11) was used to characterize the actuator. The system included: (1) a dSPACE DS1103 platform for control of algorithm implementation, (2) a DC power supply, (3) an amplifier for converting control voltage from dSPACE to current for the coils of the actuator prototype, (4) a LDV to measure the tilting (rotating) angle, and (5) the prototype with reflective tape attached at the top center to strengthen the signal. A 1 Hz sinusoidal current with amplitude ranging from 100 to 500 mA was first applied to the actuator. The direction of the current in coils on the A and B sides (or the C and D sides) (Fig. 28.8a) was made opposite so that the actuator could rotate (tilt). The frequency response of the device was also measured with a 100 mA current, ranging from 1 to 1000 Hz at the B side. In both stationary and dynamic characterization tests, a 1.25 mm by 1.25 mm square reflecting tape was cut and attached to the top of the actuating plate, with each edge parallel to the corresponding wall of the base supporter, to strengthen the signal. The laser was focused on the edge of the tape at each side to ensure that the rotating radius was 1.25 mm.

4 Experimental Results

Using the experimental setup mentioned above, the characterization results of both material and actuators are presented and discussed below.
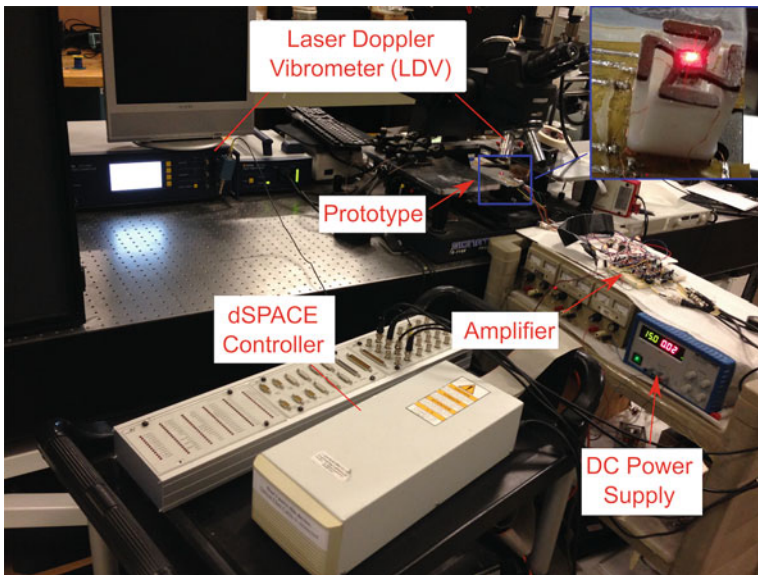
Fig. 28.11 System setup for characterizing the actuator
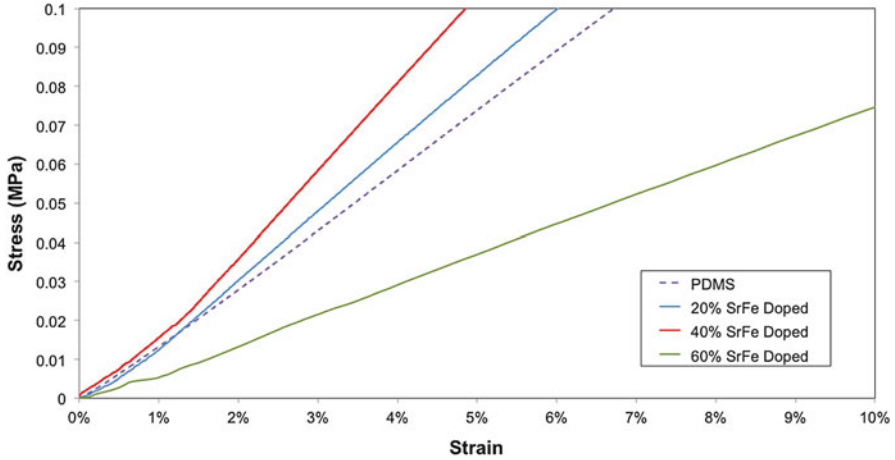
Fig. 28.12 Stress-strain curves of pure PDMS and PDMS doped with 20, 40, and 60 % w/w SrFe composites

4.1   Characterization of Material

The Young's modulus and the magnetization curve are the essential properties to be evaluated in our study. The results as well as the comparisons with the related research works are described next.

4.1.1   Young's Modulus

Figure 28.12 shows the stress-strain test results for the pure PDMS and PDMS composites with 20, 40, and 60 % wt SrFe nanopowder loaded, and Fig. 28.13 shows the elastic modulus of the samples. Young's moduli of pure PDMS is 1.53 MPa, which is consistent with [37]. For the 20 and 40 % samples, with addition of SrFe, the elastic modulus of the PDMS increased by 15.69 % and 47.71 %, respectively. Previous authors [37] implied that a reduction in Young's modulus could result in poor interactions between the PDMS matrix and addition of powders. The results suggest that for the 20 and 40 % cases, when the mixtures were stiEll liquid, SrFe and PDMS show good compatibility and SrFe may not prevent the possibility of crosslinking. For the 60 % case, the elastic moduli decreased by 47.71 %, possibly because the viscoelastic solid mixture might not have been thoroughly crosslinked, which would lead to a low modulus [5, 41].

In our case, the 60 % SrFe-doped mixture was used. A low Young's modulus of the composite is advantageous for magnetic actuation since a weaker magnetic field is needed for lower forces (less stress per unit section area), while providing the same strain (dimension change with unit length). Nevertheless, for a simple beam or plate, natural frequencies are proportional to the root of Young's modulus.
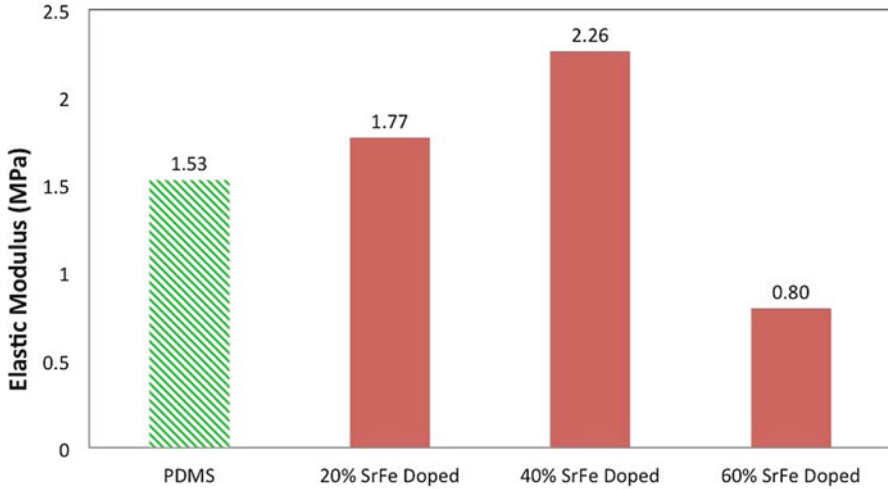
Fig. 28.13 Elastic moduli of pure PDMS and PDMS doped with 20, 40, and 60 % w/w SrFe composites

The actuating plate can be viewed as a combination of four beams and a circle plate. A lower Young's modulus means lower natural frequencies, which may not be beneficial for control in our case. Higher natural frequencies will help the system stay away from resonance.

4.1.2 Magnetization Curve

Figure 28.14 shows the M-H hysteresis curves for the samples in Table 28.1. The coercive force is consistent for all three samples, with an average value of 4.25 kOe and the remanent magnetic moment (unit weight) is proportional to the by-weight ratio of strontium ferrite-PDMS, which is also consistent with the findings of other authors [18, 28]. The coercivity is 1.85 times larger than that reported [18], which could result from magnetization during curing [12]. An external magnetic field during mixture solidification can help magnetic particles align themselves in the vertical direction. Therefore, the aligned sample will show larger coercivity, which is advantageous since the sample will be more difficult to get demagnetized. The stress-strain and M-H hysteresis results were applied to a finite element analysis (FEA) to estimate the force and magnetic field.

4.2 Dimensions

An SZ61 zoom stereo microscope (Olympus Co. Japan) was used to characterize the top view of the actuator, and ImageJ (National Institutes of Health, USA) software
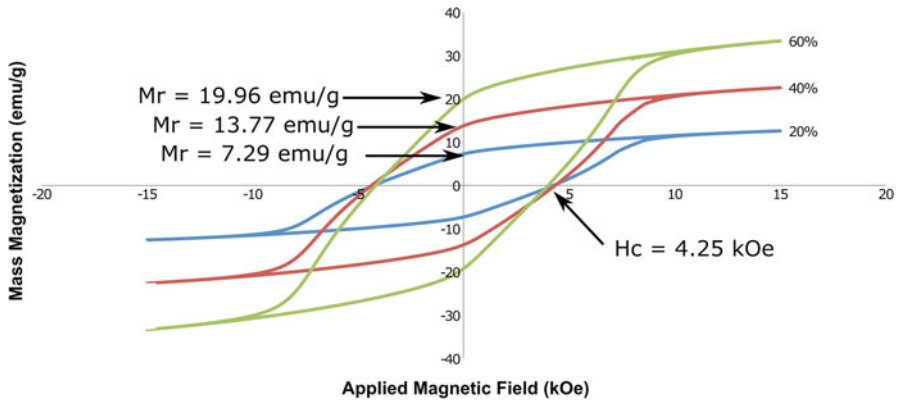
Fig. 28.14 Mass magnetization curves versus applied magnetic field for composites with 20–60 % strontium ferrite nanopowder added at 300 K (26.86 °C)

Fig. 28.15 Dimensions characterization under microscope
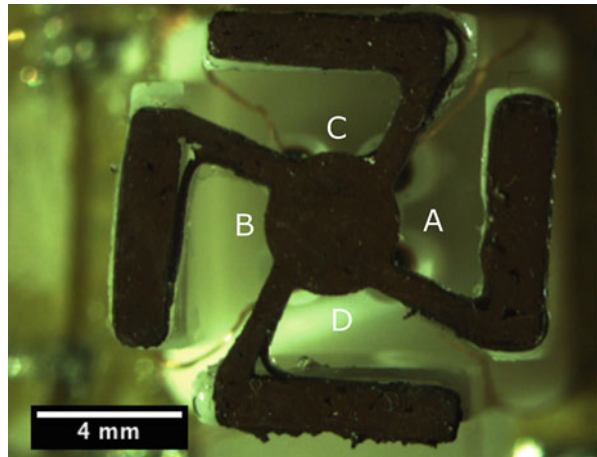
was used to calibrate and measure the dimensions (Fig. 28.15 and Table 28.2). The beams at the A, B, and C side were found to have a relatively uniform beam width, with an average width of 0.7953 mm (standard deviation of 0.00404 mm). Nevertheless, the beam at the D side was 10.6 % shorter than the average width of the others. For four beams, the average width was 0.7743 mm (standard deviation of 0.04296 mm). The variation may be due to the fact that the SLA printer generates pillars to support the printed part, and the surface of each side may not be even after the pillars are removed. We used sandpaper to file the uneven sides manually; however, some errors may have caused the asymmetric errors. In the future, the errors could be reduced by using a polishing/sanding machine instead of manual filing. The average beam length was 3.5463 mm (standard deviation of 0.1744 mm). The variation of lengths may have been the result of the assembly.

Table 28.2 Dimensions of the prototype

	Beam width (mm)	Beam length (mm)	Diameter (mm)
A	0.799	3.647	3.700
B	0.796	3.408	
C	0.791	3.740	3.862
D	0.711	3.390	

As the ring magnet and the cores of coils would attract each other, the errors in the ring magnet/plate circle and hole in the base supporter, being non-coaxial, may contribute to variations in bending even without a current. The measurements show shorter beam lengths at the B and D sides, probably because of the center of the plate shifting closer to the B and D sides. Assembling errors, along both the C and D sides, may be due to variations in the diameters (Table 28.2) and could be reduced with an optimized design for the base supporter (i.e., with grooves to hold the parallel beams).

4.3 Characterization of Actuator

To characterize the fabricated actuator, we conduct two tests, i.e., stationary test and frequency response test.

4.3.1 Stationary Test

Figures 28.16 and 28.17 show the relation between the actuating current and the tilting angle for four sides. All curves are quite linear, with R2 (coefficient of determination) larger than 0.99 and they perform a tilting rate of between 0.0011 and 0.0012 (Table 28.3), except for the C side that demonstrates a lower tilting rate. The shorter beam at the D side (Table 28.2) could cause this asymmetry. For the symmetric side, the plate is able to achieve the desired performance by being actuated by 0.6° (the maximum angle from human hand tremor [17]) with as low as a 500 mA current (50 turns/9 mm) or 1.11 mA/(mm turn), operating at a much lower current than in other reported designs [9, 27]. According to Ampere's Law [33], the magnetic flux density is proportional to the turns of the coil, current, and is reciprocal to the length. Therefore, by increasing the turns of the coil, the current used and the thickness of the actuator can be reduced. We intend to make improvements in this regard in the future.
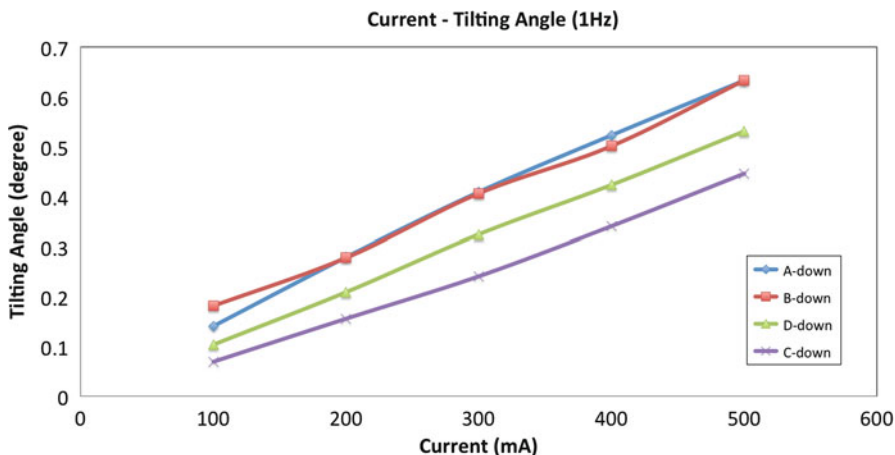
Fig. 28.16 Relation between current and tilting angle for the four sides vertically down
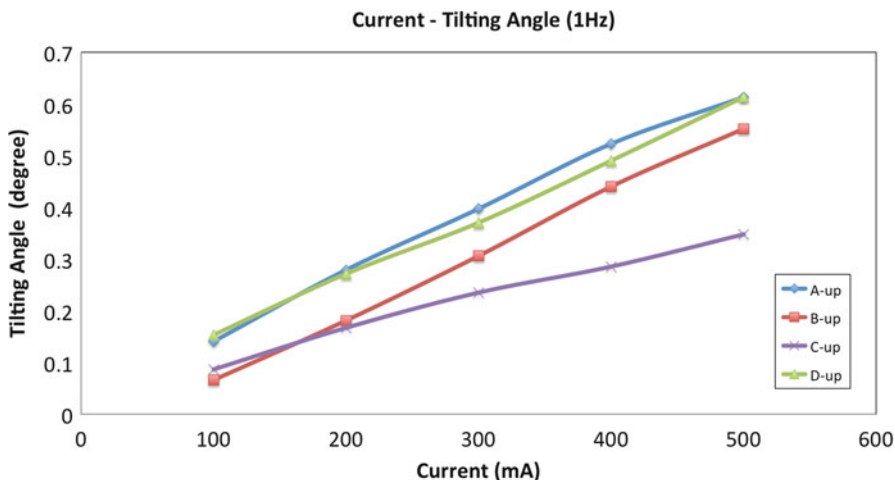
Fig. 28.17 Relation between current and tilting angle for the four sides vertically up

Table 28.3 The tilting rate for four sides with a 1 Hz sinusoidal current

Actuating side	Up tilting rate (°/mA)	Up tilting rate (°/mA)
A	0.0012	0.0012
B	0.0012	0.0011
C	0.0006	0.0009
D	0.0011	0.0011
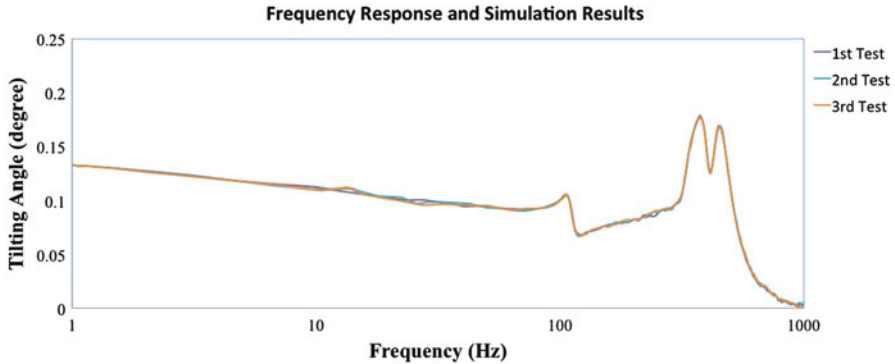
Fig. 28.18 Frequency responses of the system and simulation results

4.3.2 Frequency Response

Figure 28.18 demonstrates the frequency response for the actuator tested with the same LDV system and a digital storage oscilloscope (TDS 2014, Tektronix Inc., USA). Three tested frequency responses curves show good repeatability. The tilting angle-frequency curve shows that, at 104.11, 373.68, and 459.73 Hz, there are three peaks (natural frequencies). All normal modes are far from the operating frequency range, which is about 8–12 Hz [17] and the bandwidth is around 56 Hz, which means the ability to compensate undesired vibration under 56 Hz.

5 Uncertainty Modeling and Robust Control of Lens-Tilting OISs

To design a feedback controller for the OISs which are designed and fabricated in this chapter, we will take the following procedure. First, we sample the magnetic actuators in a batch-fabricated production line, and take their frequency responses as in Fig. 28.18. Due to limited manufacturing tolerance especially inherent to fabrication and assembly of miniature-scale devices, the frequency responses will manifest variations in resonant frequencies and DC gains. These variations can be represented as a transfer function with parametric uncertainties $G(s, \delta)$, where G is the transfer function, s is the Laplace transform variable, and δ is the parameter uncertainty vector. To the uncertain model G, we can apply well-established robust control techniques, such as robust H_∞ control [24] and μ-synthesis [35], in order to ensure stability of the closed-loop system for all OISs while optimizing the performance.

Experimental verifications of the robust control techniques on large-scale OIS prototypes with product variations have been conducted [2], while the analogous verifications with miniaturized actuator prototypes are ongoing and will be presented elsewhere.

6 Conclusion

In conclusion, we successfully applied 3D printing moulding methods to fabricate a miniature magnetic actuator for an optical image stabilizer. The Young's modulus of the magnetization curve for the PDMS composite showed an advantage in the actuating. The results of the magnetization curves are consistent with previous works. The actuator could achieve a 0.6° tilting angle with 500 mA (50 turns/9 mm), operating at a much lower current than in other designs [9, 27]. The frequency response test also showed that the actuator could easily help the system become stable and experimental results match simulation results very well, showing less than 5 % errors. In replica experiments, we analyzed the moulding errors are 0.69–10.87 % for diameter, 2.32–22.75 % for beam width, and 0.38–4.99 % for beam length. Mould fabrication errors are 2.90 % for beam width, 2.85 % error for beam length, and 1.11 % errors for diameter. We also disclosed that standard deviations of bandwidth are positive-related to errors in beam width and that variations in natural frequency between 100–1000 Hz are positive-related to errors in beam length. Finally, the application of robust control techniques to actuate the developed miniature magnetic actuators was briefly discussed. The experimental validation of robust control in miniature actuators will be a future research topic.

References

1. Ahn C, Allen M (1995) Fluid micropumps based on rotary magnetic actuators. In: Proceedings of IEEE micro electro mechanical systems workshop (MEMS'95), Amsterdam
2. Alizadegan A, Zhao P, Nagamune R, Chiao M Robust control of miniaturized optical image stabilizers against product variabilities. Submitted for publication
3. Bak D (2003) Rapid prototyping or rapid production? 3D printing processes move industry towards the latter. Assem Autom 23(4):340–345
4. Berman B (2012) 3-D printing: the new industrial revolution. Bus Horiz 55(2):155–162
5. Bokobza L, Rapoport O (2002) Reinforcement of natural rubber. J Appl Polym Sci 85(11):2301–2316
6. Bonyár A, Sántha H, Ring B, Varga M, Gábor Kovács J, Harsányi G (2010) 3D rapid prototyping technology (RPT) as a powerful tool in microfluidic development. Procedia Eng 5:291–294
7. Brian S (2015) Magnetic disk drive incorporating a magnetic actuator lock and a very small form factor. US Patent, US5034837 A
8. Chereau R, Breckon T (2013) Robust motion filtering as an enabler to video stabilization for a tele-operated mobile robot. In: SPIE security+defence. International Society for ICS and Photonics

9. Chiou J, Hung C, Lin C (2010) Design, fabrication and actuation of a MEMS-based image stabilizer for photographic cell phone applications. J Micromech Microeng 20(7):075025.1–075025.11
10. Chiu C, Chao P, Kao N, Young F (2008) Optimal design and experimental verification of a magnetically actuated optical image stabilization system for cameras in mobile phones. J Appl Phys 103(7):07F136.1–07F136.3
11. Comina G, Suska A, Filippini D (2014) PDMS lab-on-a-chip fabrication using 3D printed templates. Lab Chip 14(2):424–430
12. Cross W, Affleck L, Kuznetsov M, Parkin I, Pankhurst Q (1999) Self-propagating high-temperature synthesis of ferrites MFe2O4 (M = Mg, Ba, Co, Ni, Cu, Zn); reactions in an external magnetic field. J Mater Chem 9(10):2545–2552
13. de Silva C (2009) Modeling and control of engineering systems. CRC Press, Boca Raton
14. Dippenaar D, Schreve K (2012) 3D printed tooling for vacuum-assisted resin transfer moulding. Int J Adv Manuf Technol 64(5–8):755–767
15. Ertürk S (2002) Real-time digital image stabilization using Kalman filters. Real-Time Imaging 8(4):317–328
16. Fulcrand R, Bancaud A, Escriba C, He Q, Charlot S, Boukabache A, Gué A (2011) On chip magnetic actuator for batch-mode dynamic manipulation of magnetic particles in compact lab-on-chip. Sens Actuators B: Chem 160(1):1520–1528
17. Golik B (2006) Development of a test method for image stabilizing systems. Technical report, Department of Imaging Sciences and Media Technology, Cologne University of Applied Sciences. Diploma thesis
18. Green S, Gianchandani Y (2009) Wireless magnetoelastic monitoring of biliary stents. J Microelectromech Syst 18(1):64–78
19. Guo Y, Liu G, Xiong Y, Tian Y (2006) Study of the demolding process—implications for thermal stress, adhesion and friction control. J Micromech Microeng 17(1):9–19
20. Hao Q, Cheng X, Kang J, Jiang Y (2015) An image stabilization optical system using deformable freeform mirrors. Sensors 15(1):1736–1749
21. Hon Hai Precision Industry Co. Ltd. (2013) Plate spring for voice coil motor. US Patent, US8390154 B2
22. Hull C (1986) Apparatus for production of three-dimensional objects by stereolithography. US Patent, US4575330 A
23. Kamei K, Mashimo Y, Koyama Y, Fockenberg C, Nakashima M, Nakajima M, Li J, Chen Y (2015) 3D printing of soft lithography mold for rapid production of polydimethylsiloxane-based microfluidic devices for cell stimulation with concentration gradients. Biomed Microdev 17(2):36.1–36.8
24. Kanev S, Scherer C, Verhaegen M, de Schutter B (2004) Robust output feedback controller design via local BMI optimization. Automatica 40(7):1115–1127
25. Kauhanen P, Rouvinen J (2006) Actuator for miniature optical image stabilizer. In: Proceedings of 10th international conference on new actuators
26. Khademolhosseini F, Liu C, Lim C, Chiao M (2015) Application of periodic loads on cells from magnetic micropillar arrays impedes cellular migration. In: Proceedings of micro electro mechanical systems (MEMS), Estoril
27. Kim C, Song M, Kim Y, Park N, Park K, Park Y (2012) Design of a new triple electro-magnetic optical image stabilization actuator to compensate for hand trembling. Microsyst Technol 18(9–10):1323–1334
28. Lagorce L, Allen L (1997) Magnetic and mechanical properties of micromachined strontium ferrite/polyimide composites. J Microelectromech Syst 6(4):307–312
29. Lagorce L, Brand O, Allen M (1999) Magnetic microactuators based on polymer magnets. J Microelectromech Syst 8(1):2–9
30. Lee S, Kong J (2014) An implementation of closed-loop optical image stabilization system for mobile camera. In: Proceedings of IEEE international conference on consumer electronics (ICCE)

31. Lin C, Chiou J (2011) Design, fabrication and actuation of four-axis thermal actuating image stabiliser. Micro Nano Lett 6(7):549–552

32. Mansoor H, Zeng H, Chiao M (2011) Real-time thickness measurement of biological tissues using a microfabricated magnetically-driven lens actuator. Biomed Microdevices 13(4):641–649

33. Moskowitz BM (2015) Hitchhiker's guide to magnetism. Available online: http://www.irm.umn.edu/hg2m/hg2m_a/hg2m_a.html

34. Nisar A, Afzulpurkar N, Mahaisavariya B, Tuantranont A (2008) MEMS-based micropumps in drug delivery and biomedical applications. Sens Actuators B: Chem 130(2):917–942

35. Packard A, Doyle JC (1993) The complex structured singular value. Automatica 29(1): 71–109

36. Pallapa M, Yeow J (2013) Design, fabrication and testing of a polymer composite based hard-magnetic mirror for biomedical scanning applications. J Electrochem Soc 161(2):B3006–B3013

37. Pirmoradi F, Cheng L, Chiao M (2009) A magnetic poly(dimethylesiloxane) composite membrane incorporated with uniformly dispersed, coated iron oxide nanoparticles. J Micromech Microeng 20(1):015032.1–015032.7

38. Pournazari P, Nagamune R, Chiao M (2014) A concept of a magnetically-actuated optical image stabilizer for mobile applications. IEEE Trans Consum Electron 60(1):10–17

39. ROHM Semiconductor (2015) Optical image stabilization (OIS) white paper. Available online: http://www.rohm.com/documents/11308/12928/OIS-white-paper.pdf

40. Sachs E, Wylonis E, Allen S, Cima M, Guo H (2000) Production of injection molding tooling with conformal cooling channels using the three dimensional printing process. Polym Eng Sci 40(5):1232–1247

41. Saini D, Shenoy A, Nadkarni V (1985) Effect of surface treatments on rheological, mechanical and magnetic properties of ferrite-filled polymeric systems. Polym Eng Sci 25(13):807–811

42. Shih T, Chen C, Ho J, Chuang F (2006) Fabrication of PDMS (polydimethylsiloxane) microlens and diffuser using replica molding. Microelectron Eng 83(11–12):2499–2503

43. Song M, Hur Y, Park N, Park Y, Park K, Lim S, Park J (2009) Development of small sized actuator for optical image stabilization. In: Proceedings of international symposium on mechatronic technologies, Istanbul

44. Song M, Baek H, Park N, Park K, Yoon T, Park Y, Lim S (2010) Development of small sized actuator with compliant mechanism for optical image stabilization. IEEE Trans Magnet 46(6):2369–2372

45. Tang Y, Tan W, Fuh J, Loh H, Wong Y, Thian S, Lu L (2007) Micro-mould fabrication for a micro-gear via vacuum casting. J Mater Process Technol 192–193:334–339

46. Yu H, Liu T (2007) Design of a slim optical image stabilization actuator for mobile phone cameras. Physica Status Solidi (c) 4(12):4647–4650

47. Zhou K, Doyle JC, Glover K (1996) Robust and optimal control. Prentice Hall, New Jersey

Chapter 29
Biofeedback Technologies for Wireless Body Area Networks

Rui Li, Daniel T.H. Lai, and Wee Sit Lee

Abstract A growing trend of minimised electronic components and wireless communication technologies has brought great potential for wireless body area networks (WBAN). In WBAN, invasive or non-invasive sensors placed in on or around the human body are widely utilised to collect vital data of the human body. These signals are further analysed and processed in order to provide information on the state of the human body. Additionally, along with the development of micro-electro-mechanical-systems (MEMS), portable and wearable measurement units have become available for making this information delivery system more easily applicable in practice. Nowadays, such systems have been widely applied in medical (e.g. blood pressure measurement) and sports (e.g. movement measurement) areas. Recently, work has begun on exploring a more proactive use of WBANs, switching them from an active monitoring technology to a proactive one, namely a technology that not only monitors but also reacts autonomously to the situation. In this chapter, we focus on an emerging WBAN paradigm, namely real time biofeedback to the WBAN user. We will first review the concept of biofeedback control systems and its structure will be illustrated. Biofeedback in WBAN systems requires a multidisciplinary approach combining actuators, sensors, communications and signal processing. Existing work in this exciting new area will be highlighted before further challenges and open research issues are mentioned.

Keywords Body sensor networks • Sensors • Actuators • Biofeedback • Sports • Health

1 Introduction

In our contemporary society, the requirements of medical services arising from the aging population as well as diseases such as cardiovascular and cerebrovascular disorders are increasingly crucial. Given the population change corresponding to different age groups, it is clear that the age group of 65 and over has increased the

R. Li (✉) • D.T.H. Lai • W.S. Lee
College of Engineering and Science, Victoria University, Footscray, VIC 3011, Australia
e-mail: daniel.lai@vu.edu.au

659

Fig. 29.1 Global population
estimation by number
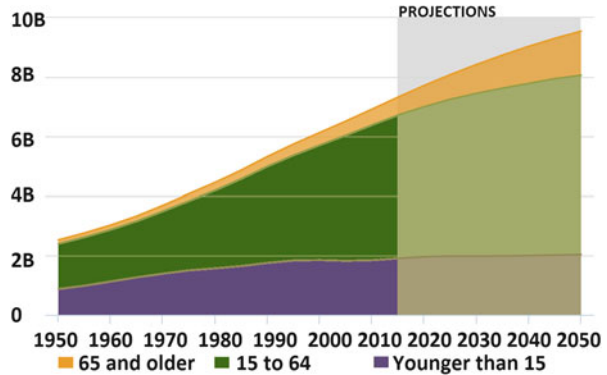
Fig. 29.2 Global population
estimation by percentage
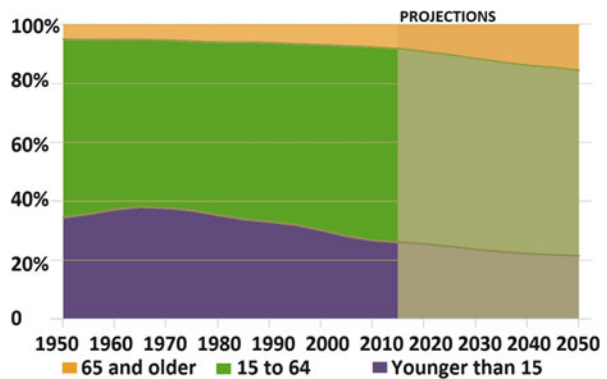
most over the last decade [1]. Compared to other age groups, this group numbers
have been steeply climbing as seen in Figs. 29.1 and 29.2 [2]. The aged population
brings significant impact on global economic development, productivity, and more
importantly, the health care system. Australia, for example, has experienced increas-
ing pressure to fund the rising costs of the national health care system over the last
decade. Statistics from the Australian Institute of Health and Welfare [3] showed the
annual national health care expenditure grew an average 5.4 % annually from 1986
to 2011, while GDP grew annually at only 3.1 % ! Within the expenditure, people
aged 65 and over cost more than $8000 per person per year, twice as much as the
cost of the younger group. In addition, top ranked causes of death such as coronary
heart disease (15 % in 2011), cerebrovascular disease (8 % in 2011), dementia (7 %
in 2011), and lung cancer (6 % in 2011)[3] have also put a toll on the healthcare
system. It has been shown that most diseases are preventable with appropriate early
interventions, and they can be well managed and treated by proactive understanding
of common risk factors [4, 5]. The consensus now is to shift the focus of healthcare
from being reactive to proactive (individuals should look after their health early
on) in order to reduce the national burden of costs and increase the efficiency
of the national healthcare system. Clearly this goal requires a dramatic change in
mentality, something that will require time. In the short term, we are forced to look
to science and engineering for possible solutions.

In sports the use of technology and innovation is fast gaining traction [6]. To date, various commercialised sports-related innovations have been brought into the stage. These include competition-level technologies such as Hawk-Eye for tracking a tennis ball's trajectory [7], Goal-Line used for validating goals in soccer games [8], advanced stopwatches used for measuring swimmers' stroke characteristics [9], as well as personal training equipment such as heart rate monitoring and global position system (GPS)-enabled devices [10]. In the meantime, research has been conducted to explore the possibility of quantifying regular training. Quantitative assessment technologies are applied to some sports such as cricket fast bowling [11], surfing [12], snowboarding [13], pistol shooting [14], cycling [15] and swimming [16, 17]. Such practical technologies are required to help understand human sensorimotor mechanisms and their interactions with machines, equipment and environment. In these applications, biofeedback is central for technologies used for sports training, injury prevention and even rehabilitation.

It is apparent that the medical and sporting needs above require more intrinsic personal monitoring, something an integrated wireless body area network (WBAN) promises to do. In typical WBANs, miniature sensors detect physiological symptoms or physical activities of human users. A central processing unit deals with data transmission, signal processing, power management and information storage [18]. The idea of adding mini actuators to WBAN platform allows one to provide feedback signals to alert users or to provide some form of augmented information. In general, biofeedback technologies will be able to spontaneously help monitor and adjust users' performance. Furthermore, it is the micro-electro-mechanical-systems (MEMS) technology which minimises the size of sensors, processors and actuators that enables the concept of WBAN highly viable and valuable. The biofeedback control systems which apply MEMS to assemble WBANs will exempt human involvement in any body area treatment or training sessions. In the meantime, it would cause minimum possible distraction and discomfort on the end users.

This chapter aims to highlight the past and current biofeedback control techniques, some of which are just beginning to be implemented in the context of WBANs. The chapter will start with the structure and basic characteristics of a feedback control system in Sect. 2, followed by Sect. 3 which focuses on various applications of biofeedback control systems including the area of motor biomechanics, health care and sports. Special requirements when applying biofeedback control systems in WBANs are explained in Sect. 4. Finally, we conclude with further challenges and open research issues in Sect. 5.

2   Biofeedback Control Systems for WBAN

Biofeedback control systems primarily consist of three elements, namely sensors, actuators and control algorithms. These elements transfer similarly when biofeedback is applied to the WBAN framework albeit with some differences as seen in this section.
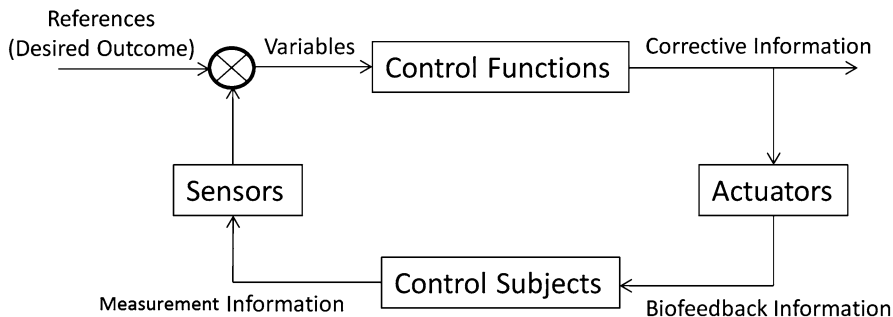
Fig. 29.3 Diagram of a biofeedback control system for WBAN

2.1 Elements in Biofeedback Control System

Structures of classical feedback control systems vary depending on control theories involved and the respective control variables. A biofeedback system for WBANs integrates the ideas of control theory by using sensor data as reference signals and actuators to provide feedback to the human body (shown in Fig. 29.3). The challenge is then to appropriately model the component of the human body of interest, e.g. locomotion, so that correct feedback can be provided.

2.1.1 Sensors

The development of sensor technology is the initiator for further design of biofeedback control systems applied in the human body area network. The sensor technologies transduce physical or chemical parameters to analog or digital signals, allowing further analysis, e.g. statistics and signal processing. Sensor readings provide the input of biofeedback control systems. Wearable sensors bring great opportunities for quantitatively understanding physical characteristics of the user. Typically, force sensors (including exoskeleton devices [19] and fingertip-mounted [20] devices) and inertial sensors (including accelerometer [21], gyroscope [22] and magnetometer [23]) are extensively applied for understanding physical activity both in static cases (e.g. ability of keeping balance [14]) and motor control tasks (e.g. posture control [17, 24–46]). Heart rate monitoring devices which are traditionally employed in medical applications have been brought to sports area in recent years. They have found more use in hazard monitoring in order to warn athletes of their physical response to the ongoing training intensity and stress [47]. Energy expenditure can be measured by direct or indirect methods such as oxygen consumption/carbon dioxide production devices [48], calorimeters [49] and heat-flow sensors [47].

Table 29.1 Biofeedback control devices (actuators)

Category		Applications
Visual feedback devices		[17, 25–32, 34, 52, 63–76]
Audio feedback devices		[50, 51, 63, 77, 78]
Tactile devices	Electro-tactile	[53, 54]
	Vibro-tactile (force-haptic)	[55–57, 79]
	Visuo-tactile (visual-haptic)	[58, 59]
	Thermal	[61, 62]

2.1.2   Actuators

Actuators are tools that can provide the feedback mechanisms to the subject to improve its current state to the desired one. Typical applications of actuators applied in human body area control systems are included in Table 29.1. Audio feedback devices are substantially used to warn control subjects of improper performance (e.g. risk postures warning [50], gait analysis and diagnosis [51]), while visual feedback mechanism emphasises the monitoring function by visually illustrating to control subjects their performance and desired objectives (e.g. 3D shoulder position measurement [52], swimmers performance measurement[17, 25–32, 34]). The third type of actuator is tactile feedback actuator devices which can be divided into four subcategories:

- Electro-tactile actuators transmit control signals through a current injection from surface electrodes into the human skin giving a prickly sensation. It requires direct contact between human skin and the electro-tactile actuators. Recently, researchers have been looking into its applications particularly in areas such as hand gesture control [53] and sway posture control [54].
- Vibro-tactile actuators are used where visual or audio information is limited. This type of actuator (usually a piezoelectric transducer) is capable of delivering various distributed stimuli with robust sensations [47]. In addition to its traditional applications such as sway warning [55] or risk indication [56], it can also be applied on a daily-basis such as being used in playing gesture control in piano training [57].
- Visual-Haptic actuators are designed according to subtle senses of humans. For instance, visual haptic feedback can be displayed as a bar chart which indicates users their electroencephalograph (EEG) activities [58]. When used together with electromyography (EMG) devices, it can provide feedback for general locomotor control improvement [59].

- Thermal actuators include, but not limited to, temperature sensors, temperature control systems (e.g. thermal-electric heating pump and water cooling system) and thermal stimulators [60]. They provide temperature information to users to assist their discrimination of different materials [61]. Additionally, the intensity of vibro-tactile feedback on human skin can be quantitatively described in accordance with thermal feedback human perceived [62]. Unfortunately, practical applications of thermal feedback mechanisms are limited due to the lack of thermal actuators developed and understanding of how to use them for biofeedback [60].

2.1.3 Control Functions

Functional control algorithms are core to control systems. The design of control algorithms should take into consideration the desired objectives, control variables and the level of control performance (e.g. accuracy, speed, etc.)[80]. Besides, there are some significant universal regulations such as the laws of physics (e.g. Kirchhoff's current and voltage law, Newton's force law, etc.)[81], science of dynamics, disturbances and uncertainties, and system homogeneity (when interconnecting multiple systems, all subsystems should be in similarly comparable performance)[80], restricting the design of control functions. Control models available range from the standard proportion-integration-differentiation (PID) control to more advanced concepts such as fuzzy control, stochastic control and model predictive control.

2.2 WBAN Architectures for Biofeedback

There are three major WBAN architectures that can be integrated with actuators for biofeedback. The selection of the architecture is critical for feedback control as it will affect the feedback intensities and the overall system lifespan. These architectures are depicted in Fig. 29.4:

- *Integrated-BAN* mode refers to the case where sensor, actuator and control hub reside on the same node.
- *Intra-BAN* mode involves a separated control centre which is used for gathering information from different sensors and sending directive feedback signals to actuators. The aim of including a single individual control hub is to reduce the complexity and power consumption of each body area node.
- *Inter-BAN* mode has an external control centre that collects sensor data and dispatches feedback signals. Since the control hub is off the body, more complex control algorithms can be implemented and thus this architecture provides the most accurate feedback information. However transmission delay times mean that the system response time is slow making this architecture more suitable for applications that do not require time critical biofeedback.

Fig. 29.4 WBAN
architectures: (1) integrated
mode structure; (2)
intra-BAN mode structure;
(3) inter-BAN mode structure
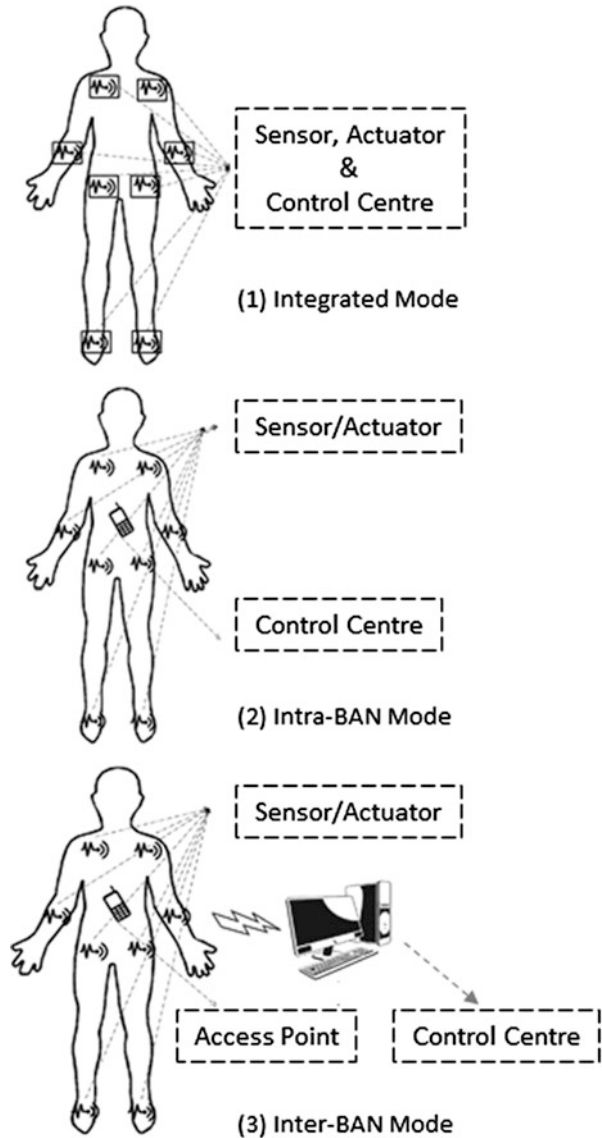
The trade-offs in architecture selection include power consumption, complexity of control algorithms, response time and control accuracy. The performance of the WBAN with respect to each trade-off parameter is included in Table 29.2.

Table 29.2 Performance of WBAN structures

	Node power consumption	Algorithm complexity	Response time	Control accuracy
Integrated mode	Highest	Lowest	Fastest	Lowest
Intra-BAN mode	Medium	Medium	Medium	Medium
Inter-BAN mode	Lowest	Highest	Slowest	Highest

3 Applications of Biofeedback Control Systems

The applications of biofeedback control systems are largely found in areas of biomechanics learning, medical therapy and sports. The key benefits of deploying such a system on human body are to provide distance control and real-time feedback to users so that appropriate treatment and prevention of hazards can be implemented effectively. The following subsections introduce some recent implementations of biofeedback control system in different fields.

3.1 Biofeedback and Human Movement

Biomechanical analysis has become the fundamental tool for understanding human locomotion. Current technologies are capable of providing extrinsic feedback to humans to assist their learning (performance improvement) and performance achievement (objective outcome). The classification of such extrinsic feedback can be divided into the following four levels (L1 or neuromuscular level, L2 or kinetic level, L3 or kinematic level and L4 or movement outcome level) based on the scope of control subjects: a piece of nerve or muscle, a body position, group of body positions and overall locomotor performance [82]. Example applications in each level are summarised in Table 29.3. It is worth mentioning that there is a growing tendency of feedback complexity along with the rise of control levels due to the increasing number of associative control variables, yet the control specificity (precision of control functions) is descending.

3.1.1 Level 1: Neuromuscular Feedback

Neuromuscular level feedback refers to the control functions specifically working on a nerve or muscle bundle. As long as three decades ago, the design of neuromuscular biofeedback system has already been developed for hemiplegia treatment [63]. Researchers found that hemiplegic patients with upper limb dysfunctions achieved certain levels of improvement in movement control after being supplied with EMG biofeedback through video and visual cues. In addition, patients showing greater potential in improving their upper extremity functions in this research could be

Table 29.3 Applications of biofeedback technologies in biomechanics and human movement

Level	Year	Objectives	Data collection	Feedback	References
L1	1983	(1) Hemiplegic patients' upper limb functions improvement (2) identify patients with higher potentials of being cured	Bioconditioner, portable EMG device	Audio and visual	[63]
L1	2003	A computer-game-based system motivates users to adhere to muscle rehabilitation programs; it improves muscle strength by comparing muscle contraction level to the predefined thresholds	EMG device	Visual feedback	[65]
L1	2012	(1) Gait parameters reading (2) a gait rehabilitation program for post-stroke patients, people suffering from cerebral palsy or other neuro-muscular gait defects, amputees, etc.	EMG devices fitted on six different muscles of the lower limb	Visual feedback	[64]
L1	2013	Motor control functions improvement for children with primary dystonia	EMG devices	Visual-haptic feedback	[59]
L1	2014	(1) A biofeedback system as an exploratory tool and a rehabilitation program for dystonia patients (2) regain senses of various physiological responses and learn to self-control	EMG devices, 6DoF Flock of Birds (FOB)	Visual-haptic feedback	[79]
L1	2014	A tele-rehabilitation program as a diagnostics tool to identify neuromuscular diseases	EMG devices	Visual feedback	[67]
L1&L3	2010	Hand rotation and grasping motion rehabilitation program for cerebral palsy (CP) and stroke victims patients for their self-training and assessing	EMG devices, magnetic motion tracking system	Virtual-reality-based visual feedback	[66]

(continued)

Table 29.3 (continued)

Level	Year	Objectives	Data collection	Feedback	References
L2	2006	A posture rehabilitation program for patients with bilateral vestibular loss	Tri-axial accelerometer	Audio feedback (lightweight headphone)	[50]
L2	2008	A step counting equipment for post stroke rehabilitation program	Force sensing resistors	Visual feedback	[78]
L2	2008	Trunk posture modification	Three uni-axis accelerometers, three uni-axial gyroscopes	Audio biofeedback (sonification of trunk kinematics)	[77]
L2	2009	A fall risk indicator which can estimate trunk tilt, decrease mediolateral sway and increase locomotor senses	Three linear accelerometers, three rate gyroscopes	Vibrotactile feedback (tactile senses replace losing senses)	[56]
L2	2011	Gait analysis for diagnosis, treatment and prevention of walking problems	F-scan pressure sensor, sole integrated gait sensing unit (6*4 matrix)	An audio warning signal or a tactile signal	[51]
L2	2013	Reduce old adults' excessive trunk sway in real time	9DOF IMU (a tri-axial accelerometer, a tri-axial gyroscope, and a tri-axial magnetometer)	Vibrotactile feedback	[55]
L3	1998	3D shoulder position measurement	6DOF electromagnetic tracking device	Visual feedback of bone and joint rotations	[52]

L3	2006	Respiratory motion detection and compensation	Electromagnetically tracked sternum sensors	Image-guided interventional procedures	[68]
L3	2010	High-precision (geometric accuracy and radiologic accuracy) radiotherapy for dynamic tumours	Phantom with three electromagnetic transponders embedded, electromagnetic array of Calypso system, accelerator	Visual feedback	[70]
L1 & L2&L3	2015	Anterior cruciate ligament (ACL) injury prevention for adolescent athletes with neuromuscular training	EMG devices, inertial sensors (accelerometers and gyroscopes), camera-based 3D motion capture	Visual feedback (augmented reality and smart-eye heads up displays)	[69]
L4	1999	(1) Feedback emphasising effect of movement rather than movement itself enhance the maintenance of balance (2) it benefits in delayed retention test without feedback provided	Stabilometer	Visual and kinesthetic feedback	[71]
L4	2002	(1) External-focus (movement effect) feedback leads to more effective learning than internal-focus (body movement) feedback does (2) reduced feedback frequency is beneficial for external feedback	Volleyball players in internal versus external feedback experiments; soccer players in feedback frequency experiments	Feedback on movement pattern (position–time curve), feedback on movement effect	[72]

particularly identified with the help of designed biofeedback system. Since then applications of neuromuscular feedback on motor function recovery treatment have been continuously studied. Later research in muscle rehabilitation programs used EMG biofeedback systems in computer-game-based routine practice for motivating patients with motion dysfunctions to complete their recovery training [65]. Similar applications in [64] have presented a virtual reality model through which patients access visual feedback on a computer screen. Here a virtual person mimicking their movement or a desired set of movements is displayed for them to learn. The virtual person was generated from fusion of data collected from six EMG devices fixed on the patients' lower limb muscles.

Lately, researchers tend to put more emphasis on the feedback mechanism and integrity of control system. Vibrotactile [79] and visual-haptic [59] feedback devices have been brought into neuromuscular feedback systems where EMG is used as the transducer device. Moreover, EMG biofeedback control system can also be organised as a diagnostic tool utilised in the tele-monitoring e-healthcare system for identifying neuromuscular diseases and managing neuro-rehabilitation progress [67].

It should be noticed that there is a developing tendency of combined applications of Level 1 and higher level biofeedback mechanisms. In [66], EMG aligned with a magnetic motion tracking device, which is categorised as the Level 3 kinematic feedback equipment, has been developed to encourage patients with upper limb disorders to have sufficient physical practice through some human–machine interactive activities. The idea of voluntary and active self-training and assessment has been raised in this research.

3.1.2   Level 2: Kinetic Feedback

Kinetic feedback mechanism is more common in human movement control systems, where the control subject normally refers to a particular position of human body, a leg or arm, for instance. Within kinetic biomechanical feedback control systems, inertial measurement units (IMUs) and force sensors are widely used sensing elements. IMUs are usually applied in cases where control subjects, older adults or patients with bilateral vestibular loss, for instance, have difficulties in controlling their posture and posing great risk of falls. Utilising audio feedback as the sensory input replacement is commonly involved in preliminary design of kinetic feedback control systems. Tri-axial accelerometer [50] or three uni-axial accelerometers plus three uni-axial gyroscopes [77] are essential in acquiring trunk tilt displacements. Wireless vibrotactile feedback (VTTF) [56] is another efficient feedback manner generated when there is an excess of the threshold of safe sway. In recent years, following with the development of sensor technology, more comprehensive and lightweight inertial measurement sensors have been developed. 9 degree-of-freedom (DOF) inertial sensor containing a tri-axial accelerometer, a tri-axial gyroscope, and a tri-axial magnetometer exerts better accuracy and adaptability in measuring gravity, orientation and angular velocity within trunk sway and posture control study [55].

Force sensors, on the other hand, is another significant tool for balance assistance and gait measurement. Force sensing resistors (FERs) can be placed in shoes for step counting [78]. Together with telecommunication protocols, they can constitute a tele-monitoring/rehabilitation program for post-stroke recovery. Similarly, F-scan pressure sensors together with audio or tactile feedback devices can form up a solo integrated gait sensor (SIGS) unit for prevention, diagnosis and treatment of walking problems [51].

3.1.3   Level 3: Kinematic Feedback

Kinematic level feedback contains integrated systematic movement information. The feedback is usually related to multiple positions of human body, so this is a dynamic and interactive feedback model. Electromagnetic tracking devices are capable of providing interventional-image-based visual feedback, which brings its applications to the kinematic level. Relevant studies conducted these years spread over various fields such as shoulder position measurement [52], respiratory motion detection [68] and tumour position detection [70].

In addition, the alliance of different level applications brings great opportunity for measurements with higher degree of complexity. For example, in [69], EMG-based muscle recruiting improvement (Level 1), force-and-inertial-sensor-based body motion detection (Level 2), camera-based 3D motion qualification (Level 3), as well as innovative visual feedback techniques have comprised an evolutionary anterior cruciate ligament (ACL) injury prevention system for young athletes [69].

3.1.4   Level 4: Movement Outcome Feedback

The highest level of biomechanical feedback control system emphasises the "knowledge" of locomotor results [82]. Feedback from this level will generally provide information on overall consequence of movement. In the study of [71, 72], the influences regarding pattern of movements and effect of movements have been investigated. The results indicate that feedback mechanism plays more profound impacts in enhancing the effect of movement than conducting the movement itself, and the learning of movement effect is more efficient than the learning of body movement. In other words, effect of movements, also known as movement outcome, can be regarded as the boost of motor performance, and this could be the ultimate objective when controlling human locomotion.

3.2   Health Care

In medical applications, biofeedback systems are widely accepted in clinical therapies. For instance, wearable respiratory biofeedback system, which includes a heart rate monitoring belt equipped with electrocardiograph (ECG) sensors and a respiration monitoring belt equipped with respiratory inductance plethysmograph

sensors, is designed for self-training and learning by letting patients follow the target respiration rhythm in a pre-defined system and watching the visual feedback on screen [73]. Robotic therapy technology for dementia care has matured in over 30 countries. MSD2.0 tele-monitoring system, a typical application, is developed to quantitatively assess patients' functional, medical, psychosocial and cognitive status [74]. Besides, EEG device is used to detect brain activities. Near-infrared spectroscopy is involved in the measurement of blood hemoglobin levels for the aim of accurately identifying active brain areas. For hypertension treatment, chest sensors are placed on patients to collect their pulses and breathing patterns. Blood pressure can be reduced if patients could well adapt to the feedback and follow the instructive inhale and exhale guidance [75]. Similarly, continuous glucose monitoring systems are developed for diabetes care. In hospitals, blood glucose level is measured by inserting a small catheter containing a subcutaneous sensor into patient body and results can be immediately displayed on monitor [76]. In home settings, wearable devices such as Glucowatch Biographer wristwatch developed by Cygnus Inc. are more convenient for conventional applications as the blood glucose is measured by an electrode/glucose oxidase/biosensor unit [76]. All the above-mentioned medical systems provide visual feedback signals to patients so that they can review their status and take responsible steps or measures. Generally, biofeedback systems have been extensively infused into clinical applications and they have shown effective functions in medical treatment and recovery therapies.

3.3 Sports

Bringing electronics technology into sports is an emerging research area where focuses are mostly placed on monitoring and improving athletes' kinematic performance, assessing athletes' psychological status and maintaining individuals' fitness in a scientific and quantitative manner. Application examples are presented in the following:

3.3.1 Athletes Kinematic Analysis

Thanks to the merits of MEMS such as the unnoticeable size, wearable sensors and technologies have been introduced into various sports fields. For instance, runners' strides can be quantitatively measured and analysed by installing a three-dimensional (3D) accelerometer providing spatial-temporal characteristics reading on runner's waist [36]. In skiing, gliding movement patterns and distance can be recognised as visual feedback by placing accelerometers (ADXL210E, Analog Devices, USA) and gyroscopes (ADXRS300, Analog Devices, USA) on skier's body, force-sensing resistors (IEE company) under athlete's feet, and infrared distance sensors (GP2D12, Sharp) on ski boot's sides [37]. Similarly, for snowboard learning, researchers have developed a mistake recognition system which applies inertial sensors (Bluetooth Shake SK6) to detect incorrect upper body rotations, two

bend sensors put at the back of each knee to detect insufficient knee flexion, and four force-sensitive resistors placed in each boot, under the ball of foot and under the heel to detect incorrect weight distribution [38]. In golf swing training, motion patterns during swing are captured by placing 3D accelerometers and 3D gyroscopes on club head, club grip, left wrist, right wrist, and at the back of waist, respectively [39].

In addition, there are some systems designed for skills acquisition and assessment. For example, position-tracking system with microwave senders installed in the football and the players' shin guards can assist game monitoring and players' performance observation [41]. Gyroscope sensors, together with optical-markers-based motion-capturing system which is also found being used in motorcycle rider's 3D position estimation [46], are employed in tennis training. Feedback of upper arm internal rotation, wrist flexion and shoulder rotation in the first tennis serve are visually displayed for coaches and biomechanists [24]. In both examples, visual feedback of athletes' performance can be set up as benchmarks for further teaching purpose or as athletes selection criterion.

Besides, there are some other type of data collection technologies designed specifically for sports area applications. In cycling practice, power meters such as embedded strain gauges installed at rear hub or on pedals are used for average power output measurement [42]. Cyclists' energy consumption can then be calculated by multiplying power they generate by time. Other techniques such as radiography [43], computed tomography [44] and magnetic resonance [45] can also be employed in aerodynamics or body position analysis because of their capabilities of supplying image feedback of cyclists' musculoskeletal tissues. All the above-mentioned biofeedback systems are deigned to provide visual feedback to users. It is also known that IMU technologies are widely applied in swimming performance monitoring and assessment. Parameters such as lap time, turn overs, stroke count, instantaneous velocity measurement and temporal phase detections have been studied in diverse literature. Typical examples of biofeedback technologies in the sports area are summarised in Table 29.4.

Experiments set-up in [40] have explored the effectiveness of different feedback contents. In this research, accelerometers were placed on rowing athletes' shoulders, hips and at the flywheel cage of the dynamic ergometer. Researchers found that detailed feedback (e.g. distance remaining, stroke rate and stroke count) is more beneficial for athletes than summary feedback (e.g. percentage scores), and no feedback case exerted worst performance.

3.3.2 Athletes Psychological Analysis

Athletes' psychological conditions have considerable impacts in their competitive levels because nervous or emotional psychological status before or during competitions will inevitably bring debilitating effects to their intrinsic performance. Date back to 1980, researchers conducted some cognitive training programs including mental rehearsal, relaxation practices and cognitive restructuring exercises to help athletes reduce their stress. Football and basketball players were invited to

Table 29.4 Micro-electro-mechanical-systems in sports

Sports	Year	Data collection	Sensor position	Design	Feedback	References
Run	2002	3D accelerometer	Waist	Spatial-temporal characteristics of runners' strides	Visual	[36]
Ski	2005	2D accelerometer, 1D gyroscope, force sensing resistors, infrared distance sensor	Skier's body, feet, ski boots	Gliding movement patterns and distance detection	Visual	[37]
Football	2005	Position tracking system	In football, in player's shin guards	Game monitoring and players' performance evaluation	Visual	[41]
Snowboard	2008	Inertial sensors (accelerometer, angular rate sensor, magnetometer, capacitive sensor, compass heading, temperature, navigation switch), bend sensors, force-sensitive resistors	Upper body, knees, in boots, under the ball of foot, under the heel	Incorrect weight distribution detection	Visual	[38]
Golf	2009	3D accelerometers, 3D gyroscopes	Club head, grip, user's wrists, back of waist	Motion patterns detect during swing	Visual	[39]
Cycle	2009	Power meter	Rear hub or pedals	Calculate cyclists' power consumption	Visual	[42]
Tennis	2009	Gyroscope, motion capture system	Upper arm, wrists, shoulders	Posture detection of first serve	Visual	[24]

				3D position estimation	Visual	[46]
Motorcycle	2013	Gyroscope, motion capture system	Rider's back	Stroke characteristics detection (stroke count, stroke rate, stroke duration, time and distance per stroke, hand entry, glide, catch, recovery)	Visual	[25, 27, 28, 30, 31, 34, 35]
Swim	2009–2013	3D accelerometer, 3D gyroscope, GPS	Head, forearm, wrist, sacrum, thigh, shank	Velocity and lap count	Visual	[26, 28–30, 35]
		3D accelerometer, 3D gyroscope	Sacrum, head	Phase detection (turn, wall push-off, propulsion)	Visual	[17, 28, 32, 33, 35]
			Wrists, forearm, sacrum	Body balance and body rotation measurement	Visual feedback through swimming goggles, tactile and acoustic feedback devices	[35]
			Upper and lower back, left and right wrists			

participate in the experiment of feedback effectiveness test. A positive result has been revealed that most players showed a reduced muscle tension and increased performance rating after being provided with EMG and heart rate feedback information [83].

Later in 2010, researchers developed more reliable biofeedback training mechanisms for athletes' peak performance training. Skiing athletes were invited to multiple biofeedback training experiments such as electrodermal response (EDR) biofeedback, EEG sensory-motor-rhythm (SMR) training and peak achievement training (PAT) [84]. The defined experiments have successfully helped those participants in terms of self-controlling their activation levels and managing their moods and emotions. Results also indicated a potential improvement in athletes' attention and concentration maintenance due to the lower anxiety level. Similar research has been launched in [85] where basketball players were supplied with heart rate variability biofeedback training (feedback on response time, concentration, heart rate variability, respiration rate and shooting differences) and motivational video training. After sufficient training sessions, stressed athletes showed an improved capability of controlling their psychophysiological status. Biofeedback training has been proved to be a promising technique to help athletes perform maximally in formal competitions.

3.3.3   Fitness Maintenance

Recently, wearable biofeedback devices have also been extensively designed and developed by many commercial industries as the assistive equipment for self-training and fitness maintenance. Biofeedback training has become an accessible and reliable routine for individuals. Typically, activity track devices assembled with GPS are used for speed and distance recording (e.g. TiMex,[1] Garmin,[2] TomTom[3]). Heart rate monitor can be added to basic activity track devices to measure exercise intensity (e.g. TICKR,[4] Megallan,[5] Fitbit,[6] SUUNTO,[7] Polar,[8] Mio Alpha,[9] Adidas[10]). Textile pressure sensors are used to improve walking or running cadences and foot landing (e.g. Sensoria Smart Sock[11]). 9-axis kinematic sensors are capable

---

[1]TiMex: http://www.timex.com/sport.

[2]Garmin: http://www.garmin.com/.

[3]TomTom: http://www.tomtom.com.

[4]TICKR: http://eu.wahoofitness.com/devices/hr.html.

[5]Megallan: http://eu.wahoofitness.com/devices/hr.html.

[6]Fitbit: http://www.fitbit.com/.

[7]SUUNTO: http://www.suunto.com/.

[8]Polar: http://www.polar.com/.

[9]Mio Alpha: http://www.mioglobal.com/Default.aspx.

[10]Adidas: http://micoach.adidas.com/smartrun/.

[11]Sensoria: http://www.sensoriafitness.com/.

of measuring steps, pace, impact G force, contact strike, and the type of foot strike. RunScribe[12] develops its data-driven athlete tracking unit by fitting a 9-axis kinematic sensor on runner's shoes and providing them with 3D analysis feedback. Power meter which measures exercise intensity in unit of watt is uniquely designed by Stryd.[13] Running Injury Clinic[14] establishes a multi-camera motion capture system for 3D gait assessment. Parvo[15] devotes to test cardiopulmonary stress by measuring indirect calorimetry and maximal oxygen consumption. All the above-mentioned technologies provide visual feedback to users which allow them to directly observe their exercise statistics or images on devices.

4 Requirements for Biofeedback Control System Applied in WBAN

Designing a biofeedback control system requires a high level of system integration, which brings design specifications of various aspects a significant effect in system performance. Figure 29.5 provides an overview of requirements in biofeedback system design.

The design and development of a biofeedback control system consists of four main tasks: system modelling, system identification, signal processing and control function integration [86]. Mathematical models are required to be capable of truly
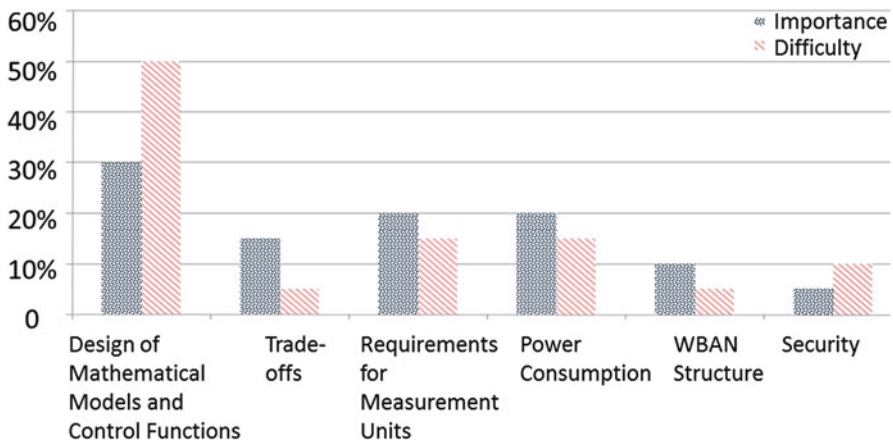
Fig. 29.5 Biofeedback control system design requirements

---

¹²RunScribe: http://www.runscribe.com/.

¹³Stryd: https://www.stryd.com/.

¹⁴Running Injury Clinic: http://runninginjuryclinic.com/clinic-services/.

¹⁵Parvo: http://www.parvo.com/.

representing real world control subjects. Human locomotion is a complex and dynamic system which requires complicated models to represent and explain its kinematic functions (e.g. static state, dynamic state, various environment impacts, etc.). When identifying an engineering control subject, system inputs, outputs, and their relationships will need to be analysed in depth in order to recognise system characteristics. Performance-relevant parameters such as noise, interferences and disturbances should be substantially accounted for. There will be various types of noise and interference collected along with the motion signals. This will emphasise a highly accurate and reliable denoising algorithm being a critical necessity. Typically, common noises collected in WBAN may because of, but not limited to, body temperature, skin moisture, limb shaking and electromagnetic radiation from other electronic devices (e.g. mobile phones). All of these factors will need to be considered in the design of control algorithms so that a feasible solution can be enabled. Besides, the choice of control functions, including the number of DOF will significantly affect the performance of designed control system. It is required to be maximally suitable for control subjects and should be decided after balancing all the trade-off parameters.

A typical trade-off parameter in the design of control system appears in the choice of an appropriate feedback gain. It is universally acknowledged that a high-gain feedback system will have more profound impacts on control subjects in terms of making decisions. Nevertheless, high feedback gain will consequently lead to the requirement of large control actions, which may be beyond the available actuation range. In this case, the high-gain solution turns to be invalid. Also, high-gain feedback may lead to the system instability resulting in self-sustaining oscillations of control subjects and increase the system sensitivity to noises [80]. Therefore, an approximate feedback gain is the essential requirement in any control systems.

It should be noticed that a suitable measurement unit is also critical in biofeedback control systems, especially when applying in WBAN. Factors such as sensor size, form factors and physical compatibility will all affect its feasibility and suitability. According to [80], requirements on measurement unit applied in WBAN are concluded as following:

- Reliability: the designed measurement unit should be capable of working within a necessary range and transmitting signals to other components of the control system.
- Accuracy: collected data should be highly accurate.
- Responsiveness: measurement units should be flexible for various changes and response quickly.
- Noise Immunity: the operation of measurement units should not be seriously affected by noises.
- Linearity: measurement units should be aware of the existence of non-linear components and try to compensate their effects.
- Non-intrusiveness: the application of measurement units should not affect the behaviour of control subjects (In WBAN case, it is a human being).

Particularly, when applying control systems on human body, it is necessary to consider the power consumption and network lifetime issues. To avoid frequent change of batteries and unexpected suspension of control functions, battery capacity should be improved and more importantly, power consumption for computation and device operation should be reduced. The requirements on power consumption are described as the balance maintenance between minimum interference, minimum tissue heating, minimum required transmission power and minimum failure [87–90]. Moreover, for the aim of reducing control function operation impacts on human bodies in terms of electronic and magnetic radiation and energy absorption by tissues, low transmission duty cycles is another requirement in addition to low transmission power consumption [90].

Additionally, security should be another issue to be considered in WBAN. According to [87], security issues include data authentication, data integrity, data confidentiality and data refreshing. All of these security aspects are set up to protect users' privacy and information's availability. There should be a security level ranking system and measurement plan for potential risks. Distinguishable management plans for different hazard priorities are expected. Also, in order to maintain a secure and constant connection between control subjects and control centre, a certain level of transmission bandwidth, throughput, delay and error rate should be maintained, even though it may consequently result in the sacrifice of power consumption. It can be another design tradeoff requiring researchers' attentions and case-based solutions.

5   Challenges in Biofeedback Control System Design

Applying biofeedback control system on human body is still in its initial concept designing stage. There are various unknown factors requiring explorations. Factors challenging further researchers in the design of biofeedback systems and their levels of significance are summarised in Table 29.5.

First of all, the mutual impact between humans and control systems remains unclear. How will humans interact with control devices (sensors/actuators), how will the control devices deal with influences such as the complex and changing environment or human body movement which may result in serious signal strength distortion, variation and attenuation [91], and how can the power consumption be minimised without compromising system stability and reliability. All of these questions are challenging for the interested researchers because very limited studies have addressed them effectively.

Secondly, when applying biofeedback control systems particularly in sports area the design of control functions normally focuses on the habituation of athletes' ability of balance maintenance or self-control of motor postures. In this process, problems of various psychological reactions, such as the capability of learning and adapting, or the diversity of feedback incentives acting on different control subjects, may be involved. Ideally, athletes are expected to adapt to feedback signals spontaneously. It means that they could react to stimulations as soon as they receive

Table 29.5 Biofeedback control systems design challenges

Factors	Level	Sub-factors	Level
Mutual interactions between humans and machines	Severe	Human vs. control devices	Major
		Control devices vs. power consumption	Moderate
		Control devices vs. changing environment and human mobility	Moderate
Control methods	Severe	Balanced merits and drawbacks	Severe
		Case-specific solutions	Moderate
Habituation	Moderate	Long-term adaption	Major
		Spontaneous reactions	Major
		Psychological reactions	Minor
Actuators	Moderate	Tactile feedback	Major
		Audio feedback	Moderate
		Visual feedback	Moderate
		Others	Minor
Interferences	Minor	Off-body Interferences	Major
		On-body Interferences	Moderate
Level of control applications	Minor	Movement outcome level	Major
		Kinematic level	Major
		Neuromuscular level	Major
		Kinetic level	Minor

them and they could react correctly. Fast and accurate reactions will dramatically descent the period of each functional control cycle and correspondingly improve the control effectiveness. However, human reaction mechanism on external feedback stimulations remains unknown. To direct a motion trajectory moving as desired trail which is calculated by control system is still challenging. From a long-term perspective, users are expected to be able to conduct self-control through a self-actuated adjustment instead of relying on machine support. To achieve this goal, long-term performance monitoring and behavioural analysis will be required.

In addition, dealing with interferences in physical layer, including on-body and off-body interferences is another challenging task. On-body interference (also known as intra-coexistence [92]) occurs when multiple sensing nodes are placed on human body and they negatively interact with each other's signal strength. Off-body interference (inter-coexistence [92]) occurs due to the application of other electronic devices within the control system operation range. Both types of interference are unavoidable. They need to be removed through the design of software functions and development of sensor technology.

Selecting an appropriate control variable or a group of co-related control variables which would assist the maintenance of body stability or the control of dynamic postures could be challenging, too. According to Sect. 3, there are four levels of

biomechanics and human movement control categorised depending on the selection of control variables, rising from bottom-layer nerve or muscle control to top-layer movement outcome control. The selection considering available complexity of control algorithms and control specificity brings flexibility for diverse control applications, yet the choice is pivotal for control performance. To choose one or some control variables that are sensitive for feedback signals and exert remarkable capability of adapting to external indications could be demanding important.

Moreover, the lack of research focusing on the selection of actuators applied to the human body requires immediate attention as the actuators critically deliver detailed feedback information. When applying same control functions to same control subjects, control performance will largely depend on control subjects' reactions to different type of actuators and different actuating intensities. In previous studies, visual feedback devices are predominantly involved. However, in some circumstances e.g., underwater sports, activities requiring strict visual attention or where visual information is limited could limit the effectiveness of visual feedback. For the case where all the available actuators could potentially work, a comparison considering factors such as impact on human performance, comfort, or minimum intrusiveness should be made for maximising the effectiveness of control systems and the biofeedback provided.

Last but not least, an appropriate control method selection is critical in WBAN design. However, the merits and drawbacks of applying them in WBAN control systems remain unclear. Due to the fact that human body control is extremely complex and the individuality of each control subject may lead to diverse reactions to a single stimulation, the universal-applied conventional solutions may be too circumscribed in this case. For instance, control methods for swimming body rotation control and for surfing standing posture control, or even control methods for recreational swimmers and for elite swimmers can all be different. This challenging characteristic of WBAN biofeedback control system brings great opportunities for discovering new and creative solutions with excellent adaptability for each human body control scenario.

6 Conclusion

In this survey, a review of the ongoing research on biofeedback system in WBAN is presented. Typical applications of biofeedback system and research merits are listed. Components included in biofeedback control system and the system structure are described. Additionally, various applicable control methods which are designed to fulfill different system requirements are provided. Finally, particular requirements when applying biofeedback system in WBAN and future research challenges are raised. Importantly, biofeedback control systems will allow for continuous monitoring and measurement of users and will be able to provide instantaneous feedback to users. In summary, this is a method which will exempt

human involvement in any body area control operations. It will realise an effective distance-control approach in WBAN, and will significantly improve the way people understand and control their movements.

References

1. Australian Bureau of Statistics (2014) Feature article: population by age and sex, Australia, States and Territories. Available: http://www.abs.gov.au/ausstats/abs@.nsf/0/1CD2B1952AFC5E7ACA257298000F2E76?OpenDocument

2. Pew Research Center (2014) Global population estimates by age, 1950–2050. Available: http://www.pewglobal.org/2014/01/30/global-population/

3. Australian Institute of Health and Welfare (2014) Australia's health 2014. Australia's health series, vol 14. Australian Institute of Health and Welfare, Canberra, pp 31–63

4. National Health Priority Action Council (NHPAC) (2006) National chronic disease strategy. In: Australian Health Ministers' Conference

5. Fulcher GR, Conner GW, Amerena JV (2004) Prevention of cardiovascular disease: an evidence-based clinical aid. Med J Aust 181:F4–F14

6. Australia Government (2010) Australian sport: the pathway to success. Commonwealth of Australia, Capital Hill

7. Hawk-Eye Innovations Ltd (2015) Available: http://www.hawkeyeinnovations.co.uk/

8. FIFA World Cup (2015) Available: http://quality.fifa.com/en/Goal-Line-Technology/

9. Autocoach (2015) Available: https://www.autocoach.com.au/

10. Polar (2015) Available: http://www.polar.com/au-en

11. Glazier PS, Wheat JS (2014) An integrated approach to the biomechanics and motor control of cricket fast bowling techniques. Sports Med 44(1):25–36

12. Paillard T, Margnes E, Portet M, Breucq A (2011) Postural ability reflects the athletic skill level of surfers. Eur J Appl Physiol 111(8):1619–1623

13. Bakken A, Bere T, Bahr R, Kristianslund E, Nordsletten L (2011) Mechanisms of injuries in world cup snowboard cross: a systematic video analysis of 19 cases. Br J Sports Med 45(16):1315–1322

14. Gianikellis K (2002) Instrumentation and measurement methods applied to biomechanical analysis and evaluation of postural stability in shooting sport. In: International research in sports biomechanics. Routledge, London

15. Zhang Y, Chen K, Yi J (2013) Rider trunk and bicycle pose estimation with fusion of force/inertial sensors. IEEE Biomed Eng 60(9):2541–2551

16. James DA, Leadbetter RI, Neeli MR, Burkett BJ, Thiel DV, Lee JB (2011) An integrated swimming monitoring system for the biomechanical analysis of swimming strokes. Sports Technol 4(3–4):141–150

17. Lee J, Leadbetter R, Ohgi Y, Thiel D, Burkett B, James DA (2011) Quantifying and assessing biomechanical differences in swim turn using wearable sensors. Sports Technol 4(3–4):128–133

18. Schmidt R, Norgall T, Mörsdorf J, Bernhard J, von der Grün T (2002) Body area network (BAN)–a key infrastructure element for patient-centered medical applications. Biomed Eng 47:365–368

19. Walairacht S, Ishii M, Koike Y (2001) Two-handed multi-fingers string-based haptic interface device. IEICE Trans Inf Syst 84(3):365–373

20. Minamizawa K, Prattichizzo D, Tachi S (2010) Simplified design of haptic display by extending one-point kinesthetic feedback to multipoint tactile feedback. In: Haptics symposium. IEEE, Waltham, pp 257–260

21. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP (2005) Accelerometer use in physical activity: best practices and research recommendations. Med Sci Sports Exerc 37(11):582–588
22. Lee KN (1994) Compass and gyroscope: integrating science and politics for the environment. Island Press, Washington
23. Campolo D, Fabris M, Cavallo G, Accoto D, Keller F, Guglielmelli E (2006) A novel procedure for in-field calibration of sourceless inertial/magnetic orientation tracking wearable devices. In: The first IEEE/RAS-EMBS international conference on biomedical robotics and biomechatronics. IEEE, Pisa, pp 471–476
24. Ahmadi A, Rowlands D, James DA (2009) Towards a wearable device for skill assessment and skill acquisition of a tennis player during the first serve. Sports Technol 2(3–4):129–136
25. James DA, Leadbetter RI, Neeli MR, Burkett BJ, Thiel DV, Lee JB (2011) An integrated swimming monitoring system for the biomechanical analysis of swimming strokes. Sports Technol 4(3–4):141–150
26. Dadashi F, Crettenand F, Millet GP, Aminian K (2012) Front-crawl instantaneous velocity estimation using a wearable inertial measurement unit. Sensors 12(10):12927–12939
27. Le Sage T, Bindel A, Conway P, Justham L, Slawson S, West A (2011) Embedded programming and real-time signal processing of swimming strokes. Sports Eng 14(1):1–14
28. Davey N, Anderson M, James DA (2008) Validation trial of an accelerometer-based sensor platform for swimming. Sports Technol 1(4–5):202–207
29. Stamm A, Thiel DV, Burkett B, James DA (2011) Towards determining absolute velocity of freestyle swimming using 3-axis accelerometers. Proc Eng 13:120–125
30. Beanland E, Main LC, Aisbett B, Gastin P, Netto K (2014) Validation of GPS and accelerometer technology in swimming. J Sci Med Sport 17(2):234–238
31. Fulton SK, Pyne DB, Burkett B (2009) Validity and reliability of kick count and rate in freestyle using inertial sensor technology. J Sports Sci 27(10):1051–1058
32. Dadashi F, Crettenand F, Millet G, Seifert L, Komar J, Aminian K (2011) Front crawl propulsive phase detection using inertial sensors. Port J Sport Sci 11:855–858
33. Dadashi F, Crettenand F, Millet GP, Seifert L, Komar J, Aminian K (2013) Automatic front-crawl temporal phase detection using adaptive filtering of inertial signals. J Sports Sci 31(11):1251–1260
34. Slawson S, Justham L, West A, Conway P, Caine M, Harrison R (2008) Accelerometer profile recognition of swimming strokes. In: The engineering of sport. Springer, Paris, pp 81–87
35. Bächlin M, Tröster, G (2012) Swimming performance and technique evaluation with wearable acceleration sensors. Pervasive Mob Comput 8(1):68–81
36. Auvinet B, Gloria E, Renault G, Barrey E (2002) Runner's stride analysis: comparison of kinematic and kinetic analyses under field conditions. Sci Sports 17(2):92–94
37. Michahelles F, Schiele B (2005) Sensing and monitoring professional skiers. IEEE Pervasive Comput 4(3):40–45
38. Spelmezan D, Borchers J (2008) Real-time snowboard training system. In: CHI'08 extended abstracts on human factors in computing systems. ACM, New York, pp 3327–3332
39. Ghasemzadeh H, Loseu V, Guenterberg E, Jafari R (2009) Sport training using body sensor networks: a statistical approach to measure wrist rotation for golf swing. In: Proceedings of the fourth international conference on body area networks. Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering (ICST), Brussels, p 2
40. Anderson R, Harrison A, Lyons GM (2005) Rowing: accelerometry-based feedback – can it improve movement consistency and performance in rowing? Sports Biomech 4(2):179–195
41. Beetz M, Kirchlechner B, Lames M (2005) Computerized real-time analysis of football games. IEEE Pervasive Comput 4(3):33–39
42. Malkinson T (2009) Current and emerging technologies in endurance athletic training and race monitoring. In: 2009 IEEE Toronto international conference on science and technology for humanity (TIC-STH). IEEE, Toronto, pp 581–586
43. Janssen I, Heymsfield SB, Wang Z, Ross R (2000) Skeletal muscle mass and distribution in 468 men and women aged 18–88 yr. J Appl Physiol 89(1):81–88

44. Augat P, Eckstein F (2008) Quantitative imaging of musculoskeletal tissue. Annu Rev Biomed Eng 10:369–390
45. Hug F, Bendahan D, Le Fur Y, Cozzone PJ, Grélot L (2004) Heterogeneity of muscle recruitment pattern during pedaling in professional road cyclists: a magnetic resonance imaging and electromyography study. Eur J Appl Physiol 92(3):334–342
46. Cheli F, Mazzoleni P, Pezzola M, Ruspini E, Zappa E (2013) Vision-based measuring system for rider's pose estimation during motorcycle riding. Mech Syst Signal Process 38(2):399–410
47. Sazonov E, Neuman MR (2014) Wearable sensors: fundamentals, implementation and applications. Elsevier, Amsterdam
48. Carter J, Jeukendrup AE (2002) Validity and reliability of three commercially available breath-by-breath respiratory systems. Eur J Appl Physiol 86(5):435–441
49. Hurkmans HL, van den Berg-Emons RJ, Stam HJ (2010) Energy expenditure in adults with cerebral palsy playing Wii sports. Arch Phys Med Rehabil 91(10):1577–1581
50. Brunelli D, Farella E, Rocchi L, Dozza M, Chiari L, Benini L (2006) Bio-feedback system for rehabilitation based on a wireless body area network. In: Fourth annual IEEE international conference on pervasive computing and communications workshops. IEEE, Pisa, pp 526–531
51. Johansson A, Shen W, Xu Y (2011) An ANT based wireless body sensor biofeedback network for medical e-healthcare. In: 7th international conference on wireless communications, networking and mobile computing. IEEE, Wuhan, pp 1–5
52. Meskers C, Vermeulen H, De Groot J, Van der Helm F, Rozing P (1998) 3D shoulder position measurements using a six-degree-of-freedom electromagnetic tracking device. Clin Biomech 13(4):280–292
53. Tamaki E, Miyaki T, Rekimoto J (2011) PossessedHand: techniques for controlling human hands using electrical muscles stimuli. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 543–552
54. Wood SJ, Black FO, MacDougall HG, Moore ST (2009) Electrotactile feedback of sway position improves postural performance during galvanic vestibular stimulation. Ann N Y Acad Sci 1164(1):492–498
55. Dehzangi O, Zhao Z, Bidmeshki M-M, Biggan J, Ray C, Jafari R (2013) The impact of vibrotactile biofeedback on the excessive walking sway and the postural control in elderly. In: Proceedings of the 4th conference on wireless health. ACM, New York, p 3
56. Wall C, Wrisley DM, Statler KD (2009) Vibrotactile tilt feedback improves dynamic gait index: a fall risk indicator in older adults. Gait Posture 30(1):16–21
57. Huang K, Do EY-L, Starner T (2008) Piano touch: a wearable haptic piano instruction system for passive learning of piano skills. In: 12th IEEE international symposium on wearable computers. IEEE, Pittsburgh, pp 41–44
58. Murguialday AR, Aggarwal V, Chatterjee A, Cho Y, Rasmussen R, O'Rourke B, Acharya S, Thakor NV (2007) Brain-computer interface for a prosthetic hand using local machine control and haptic feedback. In: IEEE 10th international conference on rehabilitation robotics. IEEE, Noordwijk, pp 609–613
59. Casellato C, Pedrocchi A, Zorzi G, Vernisse L, Ferrigno G, Nardocci N (2013) EMG-based visual-haptic biofeedback: a tool to improve motor control in children with primary dystonia. IEEE Trans Neural Syst Rehabil Eng 21(3):474–480
60. Gutierrez M, Vexo F, Thalmann D (2008) Stepping into virtual reality. Springer, New York
61. Yang G-H, Kyung K-U, Srinivasan MA, Kwon D-S (2007) Development of quantitative tactile display device to provide both pin-array-type tactile feedback and thermal feedback. In: EuroHaptics conference and symposium on haptic interfaces for virtual environment and teleoperator systems. IEEE, Tsukaba, pp 578–579
62. Yang G-H, Kwon D-S (2008) Effect of temperature in perceiving tactile stimulus using a thermo-tactile display. In: International conference on control, automation and systems (ICCAS). IEEE, Seoul, pp 266–271
63. Wolf SL, Binder-Macleod SA (1983) Electromyographic biofeedback applications to the hemiplegic patient changes in upper extremity neuromuscular and functional status. Phys Ther 63(9):1393–1403

64. Biswas K, Mazumder O, Kundu AS (2012) Multichannel fused EMG based biofeedback system with virtual reality for gait rehabilitation. In: Proceedings of the 4th international conference on intelligent human computer interaction (IHCI). IEEE, Kharagpur, pp 1–6

65. Lyons G, Sharma P, Baker M, O'Malley S, Shanahan A (2003) A computer game-based EMG biofeedback system for muscle rehabilitation. In: Proceedings of the 25th annual international conference on the engineering in medicine and biology society, vol 2. IEEE, Cancun, pp 1625–1628

66. Ma S, Varley M, Shark L-K, Richards J (2010) EMG biofeedback based VR system for hand rotation and grasping rehabilitation. In: 14th international conference on information visualisation (IV). IEEE, London, pp 479–484

67. Jafarova O, Tarasov E, Guk RY, Shtark M (2014) Experience of network neuro-rehabilitation project implementation in Russia. In: 12th international conference on actual problems of electronics instrument engineering (APEIE). IEEE, Novosibirsk, pp 491–496

68. Borgert J, Krüger S, Timinger H, Krücker J, Glossop N, Durrani A, Viswanathan A, Wood B (2006) Respiratory motion compensation with tracked internal and external sensors during CT-guided procedures. Comput Aid Surg 11(3):119–125

69. Kiefer AW, Kushner AM, et al (2015) A commentary on real-time biofeedback to augment neuromuscular training for ACL injury prevention in adolescent athletes. J Sports Sci Med 14(1):1

70. Krauss A, Nill S, Tacke M, Oelfke U (2011) Electromagnetic real-time tumor position monitoring and dynamic multileaf collimator tracking using a siemens 160 mlc: Geometric and dosimetric accuracy of an integrated system. Int J Radiat Oncol Biol Phys 79(2):579–587

71. Shea CH, Wulf G (1999) Enhancing motor learning through external-focus instructions and feedback. Hum Mov Sci 18(4):553–571

72. Wulf G, McConnel N, Gärtner M, Schwarz A (2002) Enhancing the learning of sport skills through external-focus feedback. J Mot Behav 34(2):171–182

73. Liu G-Z, Huang B-Y, Mei Z-Y, Guo Y-W, Wang L (2010) A wearable respiratory biofeedback system based on body sensor networks. In: Annual international conference of the engineering in medicine and biology society (EMBC). IEEE, Buenos Aires, pp 2497–2500

74. Shibata T (2012) Therapeutic seal robot as biofeedback medical device: qualitative and quantitative evaluations of robot therapy in dementia care. Proc IEEE 100(8):2527–2538

75. Liran M, Shalom M (2006) Biofeedback DSP based software for blood pressure decrease. In: International conference on information technology: research and education (ITRE). IEEE, Tel-Aviv, pp 184–185

76. Clarke S, Foster J (2012) A history of blood glucose meters and their role in self-monitoring of diabetes mellitus. Br J Biomed Sci 69(2):83–93

77. Giansanti D, Dozza M, Chiari L, Maccioni G, Cappello A (2009) Energetic assessment of trunk postural modifications induced by a wearable audio-biofeedback system. Med Eng Phys 31(1):48–54

78. Giansanti D, Tiberi Y, Maccioni G (2008) New wearable system for the step counting based on the codivilla-spring for daily activity monitoring in stroke rehabilitation. In: 30th annual international conference of the ieee engineering in medicine and biology society (EMBS). IEEE, Vancouver, pp 4720–4723

79. Casellato C, Maggioni S, Lunardini F, Bertucco M, Pedrocchi A, Sanger T (2014) Dystonia: altered sensorimotor control and vibro-tactile EMG-based biofeedback effects. In: XIII mediterranean conference on medical and biological engineering and computing. Springer, Cham, pp 1742–1746

80. Goodwin GC, Graebe SF, Salgado ME (2001) Control system design, vol 240. Prentice Hall, New Jersey

81. Nise NS (2007) Control systems engineering, 7th edn. Wiley, New York

82. Hong Y, Bartlett R (2008) Routledge handbook of biomechanics and human movement science. Routledge, London

83. DeWitt DJ (1980) Cognitive and biofeedback training for stress reduction with university athletes. J Sport Psychol 2(4):288–294

84. Pop-Jordanova N, Demerdzieva A, et al (2010) Biofeedback training for peak performance in sport case study. Macedonian J Med Sci 3(2):113–118
85. Paul M, Garg K, Sandhu JS (2012) Role of biofeedback in optimizing psychomotor performance in sports. Asian J Sports Med 3(1):29
86. Bao L (2000) Advanced control systems, 2nd ed. Mechanical Industrial Press ISBN 7-111-03103-2, Beijing, China, pp 4–7
87. Movassaghi S, Abolhasan M, Lipman J, Smith D, Jamalipour A (2014) Wireless body area networks: a survey. IEEE Commun Surv Tutorials 16(3):1658–1686
88. Latré B, Braem B, Moerman I, Blondia C, Demeester P (2011) A survey on wireless body area networks. Wirel Netw 17(1):1–18
89. Cavallari R, Martelli F, Rosini R, Buratti C, Verdone R (2014) A survey on wireless body area networks: technologies and design challenges. IEEE Commun Surv Tutorials 16(3):1635–1657
90. Cao H, Leung V, Chow C, Chan H (2009) Enabling technologies for wireless body area networks: a survey and outlook. IEEE Commun Mag 47(12):84–93
91. Renzo MD, Buehrer RM, Torres J (2007) Pulse shape distortion and ranging accuracy in UWB-based body area networks for full-body motion capture and gait analysis. In: IEEE Global Telecommunications Conference (GLOBECOM). IEEE, Washington, pp 3775–3780
92. Alam MM, Hamida EB, et al (2014) Advances in wearable sensor technology and its applications in mobile workforce's health monitoring and safety management. In: SPE middle east health, safety, environment & sustainable development conference and exhibition. Society of Petroleum Engineers, Doha

Chapter 30
Inverse Adaptive Controller Design for Magnetostrictive-Actuated Dynamic Systems

Zhi Li, Chun-Yi Su, and Xiuyu Zhang

Abstract Magnetostrictive actuators featuring high energy densities, large strokes, and fast responses are playing an increasingly important role in precision positioning applications. However, such actuators invariably exhibit asymmetric hysteresis nonlinearities that could cause oscillations and errors in the micro-positioning tasks. Therefore, in this chapter, an inverse adaptive controller design method is developed for the purpose of mitigating the hysteresis effect in the magnetostrictive-actuated dynamic systems. Focusing on the asymmetric hysteresis phenomenon, an asymmetric shifted Prandtl–Ishlinskii (ASPI) model and its inverse are utilized to describe and compensate the asymmetric hysteresis behaviors in the magnetostrictive actuator, respectively. To guarantee the global stability of the closed-loop system and the transient performance of the tracking error, a prescribed adaptive control method will be applied. The effectiveness of the proposed control scheme is validated on the magnetostrictive-actuated experimental platform.

Keywords Magnetostrictive actuator • Magnetostrictive-actuated dynamic systems • Nonlinear systems • Asymmetric hysteresis • ASPI model • Inverse compensator • Inverse compensation error • Prescribed adaptive control • Prescribed performance function • Backstepping control technique

1 Introduction

Magnetostrictive materials are a class of materials that change their shape when exposed to an external magnetic field. This property of the magnetostrictive materials is called magnetostriction [1] which was first discovered by James Joule

Z. Li
Eindhoven University of Technology, Eindhoven, 5600 MB, The Netherlands
e-mail: zhi.li@tue.nl

C.-Y. Su (✉)
Concordia University, Montreal, QC, Canada H3G 1M8
e-mail: cysu@alcor.concordia.ca

X. Zhang
Northeast Dianli University, Jilin, CO132012, China
e-mail: zhangxiuyu80@163.com

in 1842. Among the available magnetostrictive materials, the giant magnetostrictive material Terfenol-D is considered the most ideal material for fabricating magnetostrictive actuators. Terfenol-D is capable of providing a positive magnetostrain of 1000–2000 ppm (parts per million) at 50–200 kA/m in bulk materials. In addition, Terfenol-D shows the largest room temperature magnetostriction of any known magnetostrictive material which presents a good trade-off between high strain and high Curie temperature [2]. Featuring these properties, magnetostrictive actuators are poised to play an increasingly important role in applications of micro/nano-positioning [3], high dynamic servo valve [4], high-frequency micro-pump [5], etc.

The input and output responses of the magnetostrictive actuator are very important index to evaluate the performance of the actuator. According to the experimental tests reported in the literature [6–8], the input and output responses of the magnetostrictive actuated dynamic systems associated with different input frequencies and mechanical loads show complex nonlinear effects. Such nonlinear effects will severely deteriorate the positioning and tracking performance of the actuator and in the meantime cause inaccuracy, oscillations, and some other unexpected effects to the system [9], which poses a great challenge on applications of the actuator. The common approach for remedying the hysteresis effect is to construct a hysteresis inverse in putting in cascade as a compensator to cancel the hysteresis effect [10–12]. For constructions of the hysteresis inverse, two approaches are generally used: direct construction of complete inverse function of the hysteresis function (model) [10, 11], and use of an inverse multiplicative structure [13, 14] of the models to compensate the complicated component in the model, in which the development of the complete inverse function of the hysteresis model is not required. Direct construction of inverse function of the hysteresis model is mainly for operator based models such as Preisach model [15] and Prandtl–Ishlinskii (PI) model [10]. In both, the hysteresis is modeled by a superposition of elementary relay or play operators. However, only the PI model possesses the analytic form of the inverse [10], which explains why the PI model is becoming dominant in the direct inverse compensation approaches. In this chapter, we will focus on the direct inverse approaches owing to unknown function in hysteresis models.

It is recognized that the operator based models generally describe the symmetric hysteresis effects. However, in the magnetostrictive actuators, the hysteresis behaves asymmetrically. To keep the feature of PI model with the analytic inverse, the extension of the PI model to describe the asymmetric hysteresis behavior has been exploited in the literature, including: (1) cascading a nonlinear operator with the PI model. In [11], a modified PI model, superposition of one-sided dead-zone operators proceeded by the PI model, was proposed. This model can describe the asymmetric hysteresis behavior and has analytical solutions of its inversion, but it cannot describe the saturated hysteresis behaviors; (2) modifying the elementary play operator. In [16], the elementary play operator was redefined as right-side play operator and left-side play operator. In [17], a non-symmetric play operator was considered as the elementary operator. In [18], a generalized play operator with envelope functions was proposed, where the analytical inversion was provided with requirement of first-order derivative of the input signals.

To avoid the drawbacks of the PI extensions and preserve the advantage of the existing analytic inverse of the PI model while still being able to describe the asymmetric hysteresis behavior, an asymmetric shifted Prandtl–Ishlinskii (ASPI) model is utilized, which is constructed by three components: a PI model, a shift model, and an auxiliary function. The advantages of the proposed model are: (1) it is able to represent the asymmetric hysteresis behavior; (2) it facilitates the construction of the analytic inverse by directly utilizing the available PI inverse result in [10] without the requirement of first-order derivative of the input signals; (3) the analytical expression of the error of the inverse compensation can be derived for the asymmetric case, which will be explained as follows.

As reported in the literature, using the inverse for hysteresis compensation generally exhibits notable compensation errors, which are attributed to hysteresis characterization errors. The use of an estimated hysteresis model in deriving the model inverse would be expected to yield some degree of hysteresis compensation error. This error yields tracking error in the closed-loop control system. To accommodate such a compensation error, the analytical expression of the error of the inverse compensation is urged in the controller design. Along this line, in [19] and [20] the analytical inverse compensation error for backlash hysteresis was derived and a corresponding adaptive control scheme was then developed. For the PI model, an analytical error expression was obtained and an adaptive backstepping control scheme was developed in [21]. However, the analytical error expression was obtained only for the PI model, it has not yet been exploited for its extensions. Therefore, as listed in the third advantage, the inverse compensation error for the proposed ASPI model is analytically derived for the purpose of the controller design when the nonlinear system is preceded by the asymmetric hysteresis. In order to ensure the transient and steady-state performance of the tracking error a prescribed adaptive control scheme is employed. The developed prescribed adaptive control approach guarantees the global stability of the nonlinear system and achieves the prescribed transient and steady-state performance of the tracking error without knowledge of system parameters. To validate the developed ASPI model and the adaptive inverse hysteresis control scheme, experimental results on a magnetostrictive actuated platform are presented.

2   Magnetostrictive Actuator and Its Experimental Tests

The input and output responses of the magnetostrictive actuator are very important index to evaluate the performance of the actuator. To fully investigate the input and output responses, a magnetostrictive-actuated experimental platform is established, including a magnetostrictive actuator with mechanical loads, a capacitive sensor with a sensor driver, a power amplifier, and a dSPACE control board. The detailed information of these devices are provided as follows.

Fig. 30.1 The schematic
illustration of the
magnetostrictive actuator
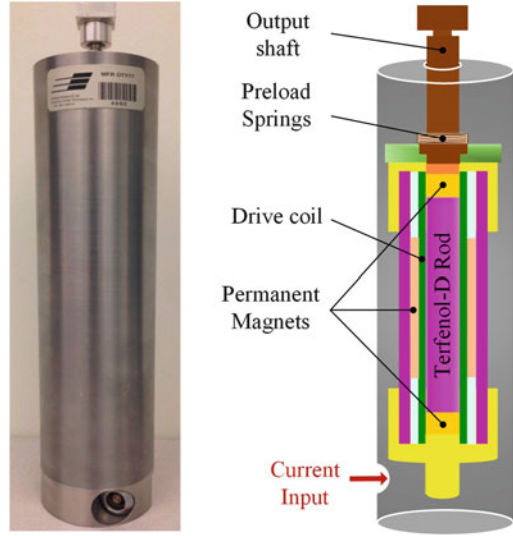
2.1 Magnetostrictive Actuators

The magnetostrictive actuators are solid state magnetic actuators and they convert
electrical current inputs into corresponding mechanical outputs. The magnetostric-
tive actuators provide an efficient way to harness the power of Terfenol-D tech-
nology and respond quickly to input current with repeatable, forceful mechanical
motion. In comparison with traditional linear electromagnetic actuators, the mag-
netostrictive actuators have the advantages of nanometric positioning resolutions,
high energy densities, large bandwidth, etc. Figure 30.1 shows the inside structure
of magnetostrictive actuator. The actuator consists of a Terfenol-D drive rod sur-
rounded by the winding coil, bias permanent magnets that produce the bidirectional
movement of the rod, a pair of preloaded springs, and an output rod attached to the
end of one Terfenol-D rod. Since the Terfenol-D rod can produce a large stroke and
output force, no additional mechanism is designed to amplify the output motion.

The magnetostrictive actuator works on the principle that when a supplied current
flows through the winding coils, a magnetic field H is created. In the presence of
the magnetic field, small magnetic domains rotate or re-orient themselves along
the magnetic lines to cause internal strains in the Terfenol-D rod. As a result, a
magnetostrictive force is produced and exerts on the output rod of the actuator,
which causes an output displacement of the output rod. The above process including
the transformation among electrical domains, magnetic domains, and mechanical
domains thus causes a very complex dynamic input and output responses.

2.2   Description of the Experimental Platform

The detailed information of the experimental platform is listed as follows:

- The magnetostrictive actuator (Model: MFR OTY77) is manufactured by Etrema Products, Inc, which provides a peak-to-peak output displacement of $100\,\mu m$ under excitations at frequencies up to 1250 Hz.
- The capacitive sensor (Lion Precision, model C23-C) with a capacitive sensor driver (Lion Precision, Elite Series CPL190) is used for measuring the displacement of the actuator with a sensitivity of $80\,mV/\mu m$, and bandwidth of 15 kHz.
- The power amplifier LVC2016 produced by AE Techron Inc. amplifies the excitation current from the dSPACE to the actuator.
- The dSPACE control board equipped with 16-bit analog-to-digital converters (ADC) and 16-bit digital-to-analog converters (DAC) is used to collect the data from the integrated capacitive sensor and apply the control signal to the amplifier. The Control Desk software in dSPACE is used for the system implementation and to interface the DS1104 PCI dSPACE board

Figure 30.2 demonstrates the magnetostrictive-actuated experimental platform. The magnetostrictive actuator is mounted on an aluminum plate via two pillow blocks. The mechanical loads are applied to the actuator via the load support frame. The capacitive sensor is fixed by an aluminum block attached to the plate. Based on the developed experimental platform, the experimental tests can be done under different input amplitudes, input frequencies as well as mechanical loads.

Figure 30.3 shows the input and output relationships of the actuated systems under different input frequencies, input amplitudes as well as applied mechanical loads. It can be observed from the experimental tests that: (1) when the magnetostrictive actuator operates with different input frequencies and without mechanical loads, the input and output responses of the actuator show dynamic hysteretic
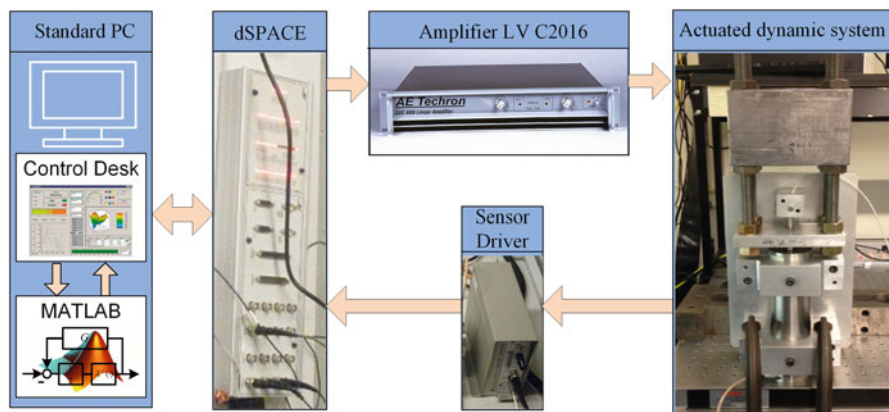
Fig. 30.2 The experimental platform with mechanical loads
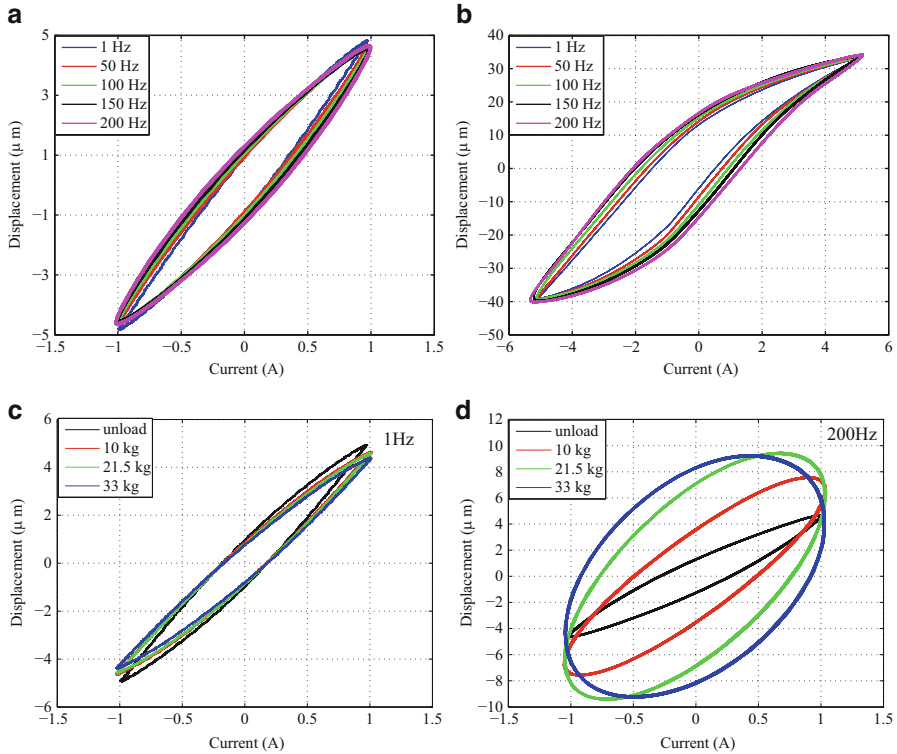
Fig. 30.3 The input and output relationship under different input frequencies, input amplitude as well as applied mechanical loads. (**a**) Different input frequencies with same input magnitude 1 A without mechanical loads. (**b**) Different input frequencies under 5 A without mechanical loads. (**c**) Different mechanical loads with same input frequency 1 Hz. (**d**) Different mechanical loads with same input frequency 200 Hz

behaviors and the width of input and output curves increase with increasing of the input frequencies; (2) when the magnetostrictive actuator operates with different input frequencies and mechanical loads, the input and output responses of the magnetostrictive actuator show strongly frequency-dependence and load-dependence properties, especially operating in high input frequency and heavy mechanical load.

From the above observation, the input and output responses of the magnetostrictive actuated dynamic system show highly nonlinear hysteresis effects, especially with mechanical loads. The existence of the hysteresis effects limits the actuating precision and performance of the actuator, and may cause undesirable inaccuracies or oscillations in the actuated systems. Therefore, in the following development, an inverse adaptive controller design method will be systematically reported for the purpose of mitigating the dynamic hysteresis effects.

3 Problem Statement

For the control purpose, the magnetostrictive-actuated dynamic systems can be expressed as a nonlinear plant preceded by a hysteresis actuator $\Pi[v](t)$, that is, the hysteresis is present as an input of the nonlinear plant as follows:

$$x^{(n)}(t) + \sum_{i=1}^{k} a_i Y_i(x(t), \dot{x}(t), \ldots, x^{(n-1)}(t)) = bu(t) \tag{30.1}$$

$$u(t) = \Pi[v](t) \tag{30.2}$$

where $v(t)$ denotes the input and $u(t)$ denotes the output, Y_i are known continuous, linear, or nonlinear functions. Parameters a_i and control gain b are unknown constants, $\Pi[v](t)$ denotes the output of the hysteresis operator, which will be described in the following development.

The control objective is to design a control signal $v(t)$ for system (30.1), such that

P1: The system state $x(t)$ tracks a desired signal $x_d(t)$ and all signals in the closed-loop are bounded;

P2: Both transient and steady-state performance of tracking error $e_1(t) = x(t) - x_d(t)$ should be within the prescribed area.

Compared with general nonlinear control for the system (30.1) only, the control signal $u(t)$ becomes the output of the hysteresis operator $u(t) = \Pi[v](t)$, where the actual control signal is $v(t)$. As it is well known, the hysteresis nonlinearity will deteriorate the system performance and cause inaccuracy or oscillations. Therefore, it imposes a challenge to handle this cascaded term with a basic requirement that $u(t)$ is not available/measurable. The common approach for remedying the effect is to construct a hysteresis inverse as a feedforward compensator. Then the control law can be designed with available control methods. The complete control scheme is shown in Fig. 30.4. It should be noted that the operator $\Pi[\cdot]$ in Fig. 30.4 is a static hysteresis operator (rate-independent). The dynamic behaviors of the magnetostrictive actuated system (rate-dependent and load-dependent phenomena in Fig. 30.3c, d) can be qualified in the formulation of the dynamic system in (30.1).

Throughout the paper the following standard assumptions are required:
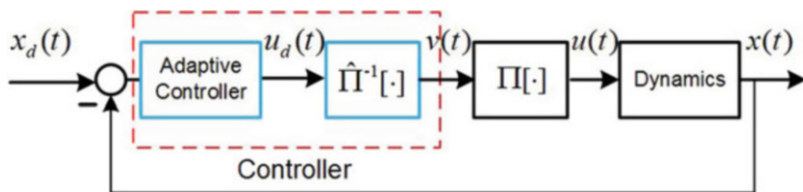
Fig. 30.4 The control scheme

Assumption 1:   The sign of uncertain parameter b in (30.1) is known. Without losing generality, it is selected as $b > 0$ in this paper.

Assumption 2:   The desired trajectory $x_d(t)$ and its (n-1)th-order derivatives are continuous. Furthermore, $[x_d, \dot{x}_d, \ldots, x_d^n]^T \in \Omega_d \subset R^{n+1}$ with Ω_d being a compact set.

4   Modeling and Inverse Construction of Asymmetric Hysteresis

In this section, the hysteresis operator $\Pi[v](t)$ defined in (30.2) will be specified as an illustration. In the literature, many hysteresis models have been proposed for representing hysteresis behaviors, such as Preisach model, Prandtl–Ishlinskii (PI) model, and Bouc–Wen model, etc. As an illustration, an extended PI model, the asymmetric shifted Prandtl–Ishlinskii (ASPI) model [22], is utilized in this chapter to describe the asymmetric hysteresis behavior in the magnetostrictive actuator, which can be seen in Fig. 30.3b. It is noted that the selection of the hysteresis model is open and interested readers may refer to [18, 23] for different asymmetric hysteresis models.

4.1   *The Asymmetric Shifted Prandtl–Ishlinskii Model*

In order to extend the PI model to describe the asymmetric hysteresis effect, still possessing its unique property of being analytically invertible, an asymmetric shifted Prandtl–Ishlinskii (ASPI) model is defined in this section, which is composed of three components: a Prandtl–Ishlinskii (PI) model, a shift model, and an auxiliary function. The purpose for introducing the shift model is to change the symmetric characteristics of the ASPI model and the auxiliary function is used for representing the saturated phenomenon. The ASPI model is thus defined as

$$u(t) \; = \; \Pi[v](t) = P[v](t) + H[v](t)$$

$$H[v] \; = \; R_1 \mathrm{sat}\left(\frac{H_1[v]}{R_1} \right)$$

$$H_1[v] \; = \; \Psi[v](t) + g(v)(t) \tag{30.3}$$

where the first term $P[v](t)$ is the PI model [10], which is defined as

$$P[v](t) = p_0 v(t) + \int_0^\infty p(r) F_r[v](t) dr \tag{30.4}$$

Fig. 30.5 Play operator

$$F_r[v]$$
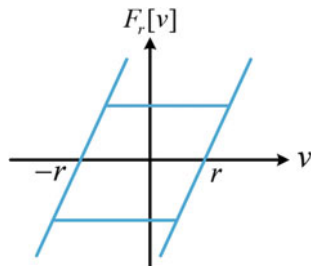
where p_0 is a positive constant; $p(r)$ is a given density function, satisfying $p(r) \geq 0$ with $\int_0^\infty rp(r)dr < \infty$. Suppose that $C_m[0, t_E]$ is the space of piecewise monotone continuous functions.

For any input $v(t) \in C_m[0, t_E]$, the play operator $F_r[v](t)$ with threshold r is defined as

$$F_r[v](0) = f_r(v(0), 0) \tag{30.5}$$

$$F_r[v](t) = f_r(v(t), F_r[v](t_j)) \tag{30.6}$$

for $t_j < t \leq t_{j+1}, 0 \leq j \leq N - 1$, with

$$f_r(v, w) = \max(v - r, \min(v + r, w)) \tag{30.7}$$

where $0 = t_0 < t_1 < \ldots < t_N = t_E$ is a partition of $[0, t_E]$, such that the function $v(t) \in C$ is monotone on each of the subintervals $[t_j, t_{j+1}]$. Figure 30.5 illustrates the input and output relationship of the play operator.

$\text{sat}(\varpi)$ is the saturation function defined as

$$\text{sat}(\varpi) = \begin{cases} +1 & \text{if } \varpi > 1 \\ \varpi & \text{if } -1 \leq \varpi \leq 1 \\ -1 & \text{if } \varpi < -1 \end{cases}$$

$R_1 > 0$ is a design parameter to define the bound of $H[v]$. $\Psi[v](t)$ is defined as the superposition of the weighted shift operators:

$$\Psi[v](t) = \int_{C_0}^{C_1} \chi(c)\Psi_c[v](t)dc \tag{30.8}$$

where $\chi(c) \geq 0$ is the density function with $\int_{C_0}^{C_1} c\chi(c)dc = L_\chi < \infty$. $\Psi_c[v](t)$ is the shift operator defined as

$$\Psi_c[v](0) = \psi_c(v(0), 0) \tag{30.9}$$

$$\Psi_c[v](t) = \psi_c(v(t), \psi_c[v](t_i)) \tag{30.10}$$

Table 30.1 Coefficients of the PI model and the ASPI model

	The PI model	The ASPI model
$p(r)$ $r \in [0, 4]$	$0.4e^{-0.01(r+\frac{1}{6})}$	$0.4e^{-0.01(r+\frac{1}{6})}$
$\chi(c)$ $c \in [1, 4]$		$0.02e^{-0.1(c-1)}$
$g(v)$		$0.8\arctan(3v - 2)$
		$-0.05v^2 + 0.05v$
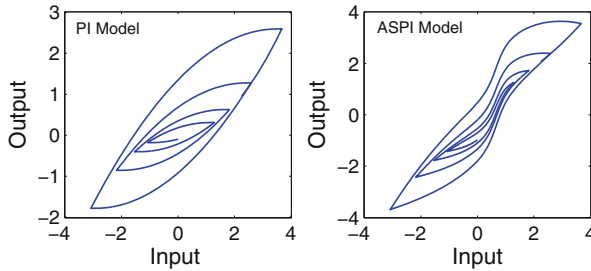
Fig. 30.6 Input–output responses of the PI model and the ASPI model

for $t_i < t \le t_{i+1}, 0 \le i \le N - 1$, with

$$\Psi_c(v, w) = \max(cv, \min(v, w)) \tag{30.11}$$

where $0 = t_0 < t_1 < \ldots < t_N$ is the partition of $[0, t_N]$. $c \in \mathbb{R}_+$, $\mathbb{R}_+ := \{x \in \mathbb{R} | x \ge 0\}$ is a parameter to determine the shape of the shift operator. When $c > 1$, $\Psi_c[v](t)$ is called left shift operator; when $0 < c < 1$, $\Psi_c[v](t)$ is called right shift operator. The last term $g(v)(t)$ in (30.3) is an auxiliary function, which assists to represent the saturation behavior of hysteresis nonlinearity. As an illustration, with a selected density function $p(r) = 0.4e^{-0.01(r+\frac{1}{6})}$, $\chi(c)$ and $g(v)$ can be correspondingly constructed, which are listed in Table 30.1 (for the detailed numerical implementation of the ASPI model, please refer to Sect. 7.1). Figure 30.6 shows responses of the PI model and the ASPI model with this selected density functions and the bound for $H[v]$ being selected as $R_1 = 10$. From Fig. 30.6, it shows the proposed ASPI model can indeed describe the asymmetric hysteresis effect.

4.2 The Inverse Construction for the ASPI Model

The challenge addressed in this section is to develop an inverse model Π^{-1} of the ASPI model for the purpose of mitigating the hysteresis effect. In fact, the inverse of Π can be constructed by utilization of the inverse result developed in [10] for the PI model itself. In the following development, we will show the procedures for this construction.

From (30.3), it is obvious that if we can find a Π^{-1} so that $\Pi[\Pi^{-1}[u]](t) = u(t)$, then such a Π^{-1} can be qualified as an inverse of the ASPI model. Since u in (30.3)

is expressed as $u(t) = \Pi[v](t) = P[v](t) + H[v](t)$, $P[v](t)$ can be re-expressed as

$$P[v](t) = u(t) - H[v](t) \tag{30.12}$$

Taking the inverse of P in (30.12) on both sides, one has

$$v(t) = P^{-1}[P[v]](t) = P^{-1}[u - H[v]](t) \tag{30.13}$$

where $P^{-1}[\cdot]$ denotes the inverse model of the PI model, which is defined as

$$P[u]^{-1}(t) = \bar{p}_0 u(t) + \int_0^{\bar{A}} \bar{p}(r) F_r[u](t) dr \tag{30.14}$$

where

$$\bar{p}_0 = \frac{1}{p_0} \tag{30.15}$$

$$\bar{p}(r) = (\varphi^{-1})''(r) \tag{30.16}$$

$$\varphi(r) = \bar{p}_0 r + \int_0^r \bar{p}(\xi)(r - \xi) d\xi \tag{30.17}$$

Thus, the following inverse Π^{-1} is obtained

$$\Pi^{-1}[u](t) = P^{-1}[u - H[v]](t) = v(t) \tag{30.18}$$

The merit of the above construction is the utilization of the analytic inverse result for the PI model in [10]. Only an extra signal $H[v](t)$ is included to the input of P^{-1} for the inverse construction of $\Pi[v](t)$.

5 Analytical Error of the Inverse Compensation for the Asymmetric Shifted Prandtl–Ishlinskii Model

It is noted that use of an estimated hysteresis model in deriving the inverse model would yield some degree of hysteresis compensation error. This error will cause tracking error in the closed-loop control system. To accommodate such a compensation error, the analytical expression of the inverse compensation error should be derived first. In [21], an analytic inverse compensation error of the PI model is developed. However, such an approach is only limited to the symmetric

hysteresis case. Focusing on the asymmetric hysteresis, in the following development we will derive the analytic inverse compensation error for the asymmetric shifted Prandtl–Ishlinskii (ASPI) model.

5.1 Overview of Composition Theorem Applied to the PI Model

In order to use the composition theorem to find the compensation error of the PI model, we first need to rewrite the PI model as [24]

$$P[u](t) = \varphi'(0)u(t) + \int_0^\Lambda \varphi''(r)F_r[u](t)dr \tag{30.19}$$

where $\varphi(r)$ denotes the initial loading curve which uniquely determines the shape of hysteresis loop described by the PI model and is defined as

$$\varphi(r) = p_0 r + \int_0^r p(\kappa)(r - \kappa)d\kappa \tag{30.20}$$

Thus, $\varphi'(0) = p_0$ is a positive constant, $\varphi''(r) = p(r)$ denotes the density function. According to the composition theorem presented in [25], the composition between two PI models $P_\gamma[\cdot](t)$ and $P_\delta[\cdot](t)$ is expressed as

$$
\begin{aligned}
P_\phi[u](t) &= P_\gamma \circ P_\delta[u](t) \\
&= \phi'(0)u(t) + \int_0^\Lambda \phi''(r)F_r[u](t)dr
\end{aligned}
\tag{30.21}
$$

where $\phi(r) = \gamma \circ \delta(r)$, $\gamma(r)$ and $\delta(r)$ denote the initial loading curves of the $P_\gamma[\cdot](t)$ and $P_\delta[\cdot](t)$, separately.

Since in practice, the exact density function $p(r)$ in the PI model may not be available. It needs to be estimated based on the measured data. In this case, the inverse model should be constructed based on the estimated density function, which is denoted as $\hat{p}(r)$. Let $\hat{P}[\cdot](t)$ denote the estimation of the actual hysteretic behavior $P[\cdot](t)$ as

$$\hat{P}[u](t) = \hat{\varphi}'(0)u(t) + \int_0^\Lambda \hat{\varphi}''(r)F_r[u](t)dr \tag{30.22}$$

where $\hat{\varphi}(r)$ is defined as

$$\hat{\varphi}(r) = \hat{p}_0 r + \int_0^r \hat{p}(\kappa)(r - \kappa)d\kappa \tag{30.23}$$

$\hat{P}^{-1}[\cdot](t)$ denotes the inverse of $\hat{P}[\cdot](t)$ as

$$\hat{P}[u]^{-1}(t) = \hat{\hat{p}}_0 u(t) + \int_0^{\bar{\Lambda}} \hat{\hat{p}}(r) F_r[u](t) dr \tag{30.24}$$

where,

$$\hat{\hat{p}}_0 = \frac{1}{\hat{p}_0} \tag{30.25}$$

$$\hat{\hat{p}}(r) = (\hat{\varphi}^{-1})''(r) \tag{30.26}$$

$$\hat{\varphi}^{-1}(r) = \hat{\hat{p}}_0 r + \int_0^r \hat{\hat{p}}(\xi)(r - \xi) d\xi \tag{30.27}$$

Thus, by applying the composition theorem on the $P[\cdot](t)$ and $\hat{P}^{-1}[\cdot](t)$ yields

$$
\begin{aligned}
u(t) &= P \circ \hat{P}^{-1}[u_d](t) \\
&= \phi'(0)u_d(t) + \int_0^{\Lambda} \phi''(r) F_r[u_d](t) dr \tag{30.28}
\end{aligned}
$$

where u_d is the desired input signal. The compensation error $e_{pi}(t)$ can be analytically written as

$$
\begin{aligned}
e_{pi}(t) &= u_d(t) - u(t) \\
&= (1 - \phi'(0))u_d(t) - \int_0^{\Lambda} \phi''(r) F_r[u_d](t) dr \tag{30.29}
\end{aligned}
$$

5.2   Analytical Error of the Inverse Compensation for the Asymmetric Shifted Prandtl–Ishlinskii Model

Due to the presence of the estimation error, we use $\hat{\Pi}[u](t)$ to estimate the true hysteresis phenomenon $\Pi[u](t)$, which is expressed as

$$\hat{\Pi}[v](t) = \hat{P}[v](t) + \hat{H}[v](t) \tag{30.30}$$

where $\hat{H}[v] = \hat{\Psi}[v](t) + \hat{g}(v)(t)$.

The output of the composition between the inverse compensation $\hat{\Pi}^{-1}[u](t)$ and true hysteretic behavior $\Pi[u](t)$ is expressed as

$$u(t) = \Pi \circ \hat{\Pi}^{-1}[u_d](t) = P \circ \hat{P}^{-1}[u_d - \hat{H}[v]](t) + H[v](t) \tag{30.31}$$

According to the combined results in (30.21), (30.31) becomes

$$u(t) = \phi'(0)(u_d - \hat{H}[v](t)) + \int_0^\Lambda \phi''(r)F_r[u_d - \hat{H}[v]](t)dr + H[v](t)$$

(30.32)

Because of $E_r[v](t) + F_r[v](t) = v(t)$, where $E_r[v](t)$ denotes the stop operator as

$$E_r[v](0) = e_r(v(0) - w_{-1}) \tag{30.33}$$

$$E_r[v](t) = e_r(v(t) - v(t_i) + E_r[v](t_i)) \tag{30.34}$$

for $t_i < t \le t_{i+1}$ and $0 \le i \le N - 1$, with

$$e_r(v) = \min(r, \max(-r, v)) \tag{30.35}$$

w_{-1} is the initial value. Then, we have

$$
\begin{aligned}
u(t) &= \phi'(0)(u_d - \hat{H}[v](t)) + \int_0^\Lambda \phi''(r)(u_d - \hat{H}[v](t))dr \\
&\quad - \int_0^\Lambda \phi''(r)E_r[u_d - \hat{H}[v]](t)dr + H[v](t) \\
&= \phi'(0)(u_d - \hat{H}[v](t)) + (\phi'(\Lambda) - \phi'(0))(u_d - \hat{H}[v](t)) \\
&\quad - \int_0^\Lambda \phi''(r)E_r[u_d - \hat{H}[v]](t)dr + H[v](t) \\
&= \phi'(\Lambda)u_d - d_b(t)
\end{aligned}
$$

(30.36)

where $d_b(t) = \phi'(\Lambda)\hat{H}[v](t) - H[v](t) + \int_0^\Lambda \phi''(r)E_r[u_d - \hat{H}[v]](t)dr$. The estimation (inverse compensation) error $e(t)$ of the ASPI model is therefore expressed as

$$e(t) = u_d(t) - u(t) = (1 - \phi'(\Lambda))u_d(t) + d_b(t) \tag{30.37}$$

It should be noted that if the estimated hysteresis operator $\hat{\Pi}[\cdot](t)$ is equal to the true hysteresis $\Pi[\cdot](t)$, it yields $\phi(r) = r$, $\phi'(r) = 1$, $\phi''(r) = 0$, then in (30.37) $\phi'(\Lambda) = 1$, $d_b(t) = 0$, leading to $e(t) = 0$. Before showing the way to utilize the estimation error in the next section, the following lemma is exploited to facilitate the robust controller design.

Lemma *The term $d_b(t)$ in (30.37) is bounded, i.e. $|d_b(t)| \le D_1$ where D_1 is a bounded constant, for any time $t \ge 0$.*

Proof Based on the definition of the stop operator [24], one has

$$|E_r[\cdot](t)dr| \le r \le \Lambda \tag{30.38}$$

From (30.38), we have

$$\int_0^\Lambda \phi''(r)E_r[u_d - \hat{H}[v]](t)dr \leq \Lambda \int_0^\Lambda \phi''(r)dr$$
$$\leq \Lambda(\phi'(\Lambda) - \phi'(0)) \qquad (30.39)$$

Thus,

$$\left| \int_0^\Lambda \phi''(r)E_r[u_d - \hat{H}[v]](t)dr \right| \leq |\Lambda(\phi'(\Lambda) - \phi'(0))| \qquad (30.40)$$

Here, we slightly modify the $H[v]$ term in the ASPI model as

$$\hat{H}[v] = R_1 \text{sat}\left(\frac{\hat{H}_1[v]}{R_1}\right)$$
$$\hat{H}_1[v] = \hat{\Psi}[v](t) + \hat{g}(v)(t) \qquad (30.41)$$

Since $H[v]$ is a designed term, we just put a bound to the function. Moreover, if the output of $\hat{H}_1[v]$ is less than the bound R_1, $\hat{H}[v] = \hat{H}_1[v]$, and normally R_1 can be set as a sufficiently large value. According to the definition in (30.41), it yields

$$-R_1 \leq H[v] \leq R_1 \qquad (30.42)$$
$$-\phi'(\Lambda)R_1 \leq \phi'(\Lambda)\hat{H}[v] \leq \phi'(\Lambda)R_1 \qquad (30.43)$$

From (30.42) and (30.43), we have

$$|\phi'(\Lambda)\hat{H}[v] - H[v]| \leq |(\phi'(\Lambda) + 1)R_1| \qquad (30.44)$$

Therefore, based on (30.40) and (30.44),

$$|d_b(t)| \leq |(\phi'(\Lambda) + 1)R_1| + |\Lambda(\phi'(\Lambda) - \phi'(0))| = D_1 \qquad (30.45)$$

6 Prescribed Adaptive Control

Different from the standard procedure of backstepping control presented in the literature [9, 26], the transient and steady-state performance of tracking error are incorporated in the design procedure of prescribed adaptive control. This control approach is originally developed in [27], which is the first time that provides a systematic procedure to accurately compute the required bounds, thus making tracking error converge to a predefined arbitrarily small residual set, with convergence rate no less than a pre-specified value, exhibiting a maximum overshoot less than a sufficiently small preassigned constant [27, 28].

6.1 Prescribed Performance Function and Error Transformation

The performance function is introduced in [27] for the purpose of depicting a convergent zone in which the trajectory of tracking error which starts from a point in the zone remains for all future time. The performance function is a decreasing smooth function, which is defined as $\rho : R^+ \to R^+$ with $\lim_{t\to\infty} \rho(t) = \rho_\infty > 0$.

It is noted that the control objective P2 can be guaranteed by satisfying

$$\underline{M}\rho(t) < e_1(t) < \overline{M}\rho(t) \tag{30.46}$$

for all $t \geq 0$, where $\underline{M} < 0, \overline{M} > 0$ are selected parameters. $\overline{M}\rho(0)$ and $\underline{M}\rho(0)$ represent the upper bound of the maximum overshoot and the lower bound of the undershoot. The constant ρ_∞ denotes the maximum tracking error at the steady state. Thus, the performance function and the parameters $\overline{M}$, $\underline{M}$ prescribe the convergent zone for the transient and steady-state performance of the tracking error.

In order to meet the requirements P1 and P2 together with condition (30.46), an error transformation is developed [27] by transforming the original nonlinear system (30.1) into an equivalent unconstraint one. Define $S(\cdot)$ a smooth and strictly increasing function and z_1 a transformed error as

$$e_1(t) = \rho(t)S(z_1) \tag{30.47}$$

$S(\cdot)$ conforms the following conditions:

1. $\underline{M} < S(z_1) < \overline{M}$
2. $\lim_{z_1\to+\infty} S(z_1) = \overline{M}, \lim_{z_1\to-\infty} S(z_1) = \underline{M}$

Since $S(\cdot)$ is strictly increasing as well as $\rho(t) > 0$, the inverse transformation can be written as

$$z_1 = S^{-1}\left(\frac{e_1(t)}{\rho(t)}\right) \tag{30.48}$$

Assume $z_1(t)$ remains bounded $z_1 \in L_\infty, \forall t \geq 0$, then $\underline{M} < S(z_1) < \overline{M}$ holds, and hence the condition (30.46) can be guaranteed. A candidate function $S(\cdot)$ is selected as

$$S(z_1) = \frac{\overline{M}e^{z_1} + \underline{M}e^{-z_1}}{e^{z_1} + e^{-z_1}} \tag{30.49}$$

Conduct inverse transformation on (30.49), yielding

$$z_1 = S^{-1}\left(\frac{e_1(t)}{\rho(t)}\right) = \frac{1}{2}ln\frac{e_1(t)/\rho(t) - \underline{M}}{\overline{M} - e_1(t)/\rho(t)} \tag{30.50}$$

Then the derivative of z_1 with respect to time can be written as

$$
\begin{aligned}
\dot{z}_1 &= \frac{\partial S^{-1}}{\partial \frac{e_1(t)}{\rho(t)}} \left(\frac{e_1(t)}{\rho(t)} \right) \\
&= \frac{1}{2} \left[\frac{1}{e_1(t)/\rho(t) - \underline{M}} - \frac{1}{e_1(t)/\rho(t) - \overline{M}} \right] \left(\frac{\dot{e}_1(t)}{\rho(t)} - \frac{e_1(t)\dot{\rho}(t)}{\rho^2(t)} \right) \\
&= r_1(\dot{x}_1 - \dot{x}_d - e_1(t)\dot{\rho}(t)/\rho(t))
\end{aligned}
\tag{30.51}
$$

where $r_1 = \frac{1}{2\rho(t)} \left[\frac{1}{e_1(t)/\rho(t) - \underline{M}} - \frac{1}{e_1(t)/\rho(t) - \overline{M}} \right]$. It is noted that both $e_1(t)$ and $\rho(t)$ in (30.51) are available and they can be involved in controller design.

6.2 Prescribed Adaptive Controller Design

The system (30.1) can be rewritten as

$$
\begin{aligned}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= x_3 \\
&\vdots \\
\dot{x}_{n-1} &= x_n \\
\dot{x}_n &= \boldsymbol{a}^T \boldsymbol{Y} + bu(t)
\end{aligned}
\tag{30.52}
$$

where $\boldsymbol{a} = [-a_1, -a_2, \cdots, -a_k]^T$ and $\boldsymbol{Y} = [Y_1, Y_2, \cdots, Y_k]^T$. The parameters $\boldsymbol{a}$, b are unknown. $u(t)$ denotes the system input with the inverse compensation as

$$
u(t) = \phi'(\Lambda)u_d - d_b(t)
\tag{30.53}
$$

Considering the time derivative of transformed error (30.51) and nonlinear system (30.52), the transformed nonlinear system dynamics are given by

$$
\begin{aligned}
\dot{z}_1 &= r_1(x_2 - \dot{x}_d - e_1(t)\dot{\rho}(t)/\rho(t)) \\
\dot{x}_2 &= x_3 \\
&\vdots \\
\dot{x}_{n-1} &= x_n \\
\dot{x}_n &= \boldsymbol{a}^T \boldsymbol{Y} + b_p u_d(t) - d(t)
\end{aligned}
\tag{30.54}
$$
$$
\tag{30.55}
$$

where $b_p = b\phi'(\Lambda)$, $d(t) = bd_b(t)$. Thus, the entire transformed dynamic system can be further written as

$$
\begin{aligned}
\dot{z}_1 &= r(x_2 - \dot{x}_d - e_1(t)\dot{\rho}(t)/\rho(t)) \\
\dot{z}_2 &= x_3 - \ddot{x}_d - \dot{\alpha}_1 \\
&\ \ \vdots \\
\dot{z}_{n-1} &= x_n - x_d^{n-1} - \dot{\alpha}_{n-2} \\
\dot{z}_n &= \mathbf{a}^T \mathbf{Y} + b_p u_d(t) - d(t) - x_d^n - \dot{\alpha}_{n-1}
\end{aligned}
\tag{30.56}
$$

The controller design is achieved by using the recursive back-stepping technique and is summarized as follows. The control law is developed as

$$
u_d(t) = \hat{\zeta} u_{d1}(t)
\tag{30.57}
$$

with

$$
u_{d1}(t) = -k_n z_n - z_{n-1} + \dot{\alpha}_{n-1} + x_d^{(n)} - \hat{\mathbf{a}}^T \mathbf{Y} + \text{sgn}(z_n)\hat{D}
\tag{30.58}
$$

where

$$
z_1 = \frac{1}{2}\ln\frac{e_1(t)/\rho(t) - M}{M - e_1(t)/\rho(t)}
\tag{30.59}
$$

$$
z_i = x_i - x_d^{(i-1)} - \alpha_{(i-1)}, i = 2, 3, \ldots, n
\tag{30.60}
$$

$$
\alpha_1 = -k_1 z_1/r_1 + e_1(t)\dot{\rho}(t)/\rho(t)
\tag{30.61}
$$

$$
\alpha_2 = -k_2 z_2 + \dot{\alpha}_1 - r_1 z_1
\tag{30.62}
$$

$$
\alpha_i = -k_i z_i + \dot{\alpha}_{i-1} - z_{i-1}
\tag{30.63}
$$

where k_i are positive designed parameters. The parameters $\hat{\zeta}$, $\hat{D}$ and the vector $\hat{\mathbf{a}}$ are updated by the following adaptation laws:

$$
\dot{\hat{\zeta}} = -\eta_\zeta u_{d1}(t) z_n
\tag{30.64}
$$

$$
\dot{\hat{\mathbf{a}}} = \mathbf{\Gamma}_a \mathbf{Y} z_n
\tag{30.65}
$$

$$
\dot{\hat{D}} = -\eta_D|z_n|
\tag{30.66}
$$

where $D = bD_1$ and D_1 is the bound defined in the Lemma in Sect. 5.2. The stability of the closed-loop system is established in the following theorem.

Theorem *For the transformed nonlinear system* (30.56) *preceded by ASPI model and its inverse compensator, the prescribed adaptive controller presented by* (30.57)–(30.66) *guarantees that*

1. *All signals in the closed-loop system remain bounded;*
2. *The tracking control with prescribed performance condition* (30.46) *is preserved.*

Proof From (30.51), and (30.59)–(30.63), and with $b_p u_d(t) = b_p \hat{\zeta} u_{d1} = u_{d1} - b_p \tilde{\zeta} u_{d1}$, we have

$$z_1 \dot{z}_1 = r z_1 z_2 - k_1 z_1^2 \tag{30.67}$$

$$z_2 \dot{z}_2 = z_2 z_3 - k_2 z_2^2 - r z_1 z_2 \tag{30.68}$$

$$z_i \dot{z}_i = z_i z_{i+1} - k_i z_i^2 - z_{i-1} z_i \tag{30.69}$$

$$z_n \dot{z}_n = z_n (-k_n z_n - z_{n-1} + \tilde{\boldsymbol{a}}^T \boldsymbol{Y} + \mathrm{sgn}(z_n) \hat{D}$$
$$\qquad -d(t) - b_p \tilde{\zeta} u_{d1}) \tag{30.70}$$

where $\tilde{\zeta} = \zeta - \hat{\zeta}$, $\tilde{\boldsymbol{a}} = \boldsymbol{a} - \hat{\boldsymbol{a}}$. Let $\tilde{D} = D - \hat{D}$. To establish the global boundedness, the following Lyapunov function candidate is adopted:

$$V(t) = \sum_{i=1}^{n} \frac{1}{2} z_i^2 + \frac{1}{2} \tilde{\boldsymbol{a}}^T \boldsymbol{\Gamma}_a^{-1} \tilde{\boldsymbol{a}} + \frac{b_p}{2\eta_\zeta} \tilde{\zeta}^2 + \frac{1}{2\eta_D} \tilde{D}^2 \tag{30.71}$$

The derivative of $V(t)$ with regard to the time is

$$\dot{V}(t) = -\sum_{i=1}^{n} k_i z_i^2 + \tilde{\boldsymbol{a}}^T \boldsymbol{Y} z_n - b_p \tilde{\zeta} u_{d1} z_n + \mathrm{sgn}(z_n) \hat{D} z_n$$

$$\qquad -d(t) z_n + \tilde{\boldsymbol{a}}^T \boldsymbol{\Gamma}_a^{-1} \dot{\tilde{\boldsymbol{a}}} + \frac{b_p}{\eta_\zeta} \dot{\tilde{\zeta}} \tilde{\zeta} + \frac{1}{\eta_D} \dot{\tilde{D}} \tilde{D}$$

$$\leq -\sum_{i=1}^{n} k_i z_i^2 + \tilde{\boldsymbol{a}}^T (\boldsymbol{Y} z_n + \boldsymbol{\Gamma}_a^{-1} \dot{\tilde{\boldsymbol{a}}}) - b_p \tilde{\zeta} (u_{d1} z_n$$

$$\qquad -\frac{1}{\eta_\zeta} \dot{\tilde{\zeta}}) - \tilde{D} (|z_n| - \frac{1}{\eta_D} \dot{\tilde{D}})$$

$$= -\sum_{i=1}^{n} k_i z_i^2 \tag{30.72}$$

Equations (30.71) and (30.72) show that $V(t)$ is nonincreasing. Therefore, $z_i (i = 1, \ldots, n)$, $\hat{\zeta}$, $\hat{\boldsymbol{a}}$, and $\hat{D}$ are bounded. By utilizing the Lasalle–Yoshizawa theorem in [29] to (30.72), it further follows that $z_i \to 0 (i = 1, \ldots, n)$ as $t \to \infty$, which concludes the tracking error is bounded within the prescribed zone.

7 Experimental Results

In this section, the prescribed adaptive controller designed above will be verified via the magnetostrictive-actuated dynamic system. The inverse compensator is first constructed and implemented in the dSPACE. Then the prescribed adaptive controller is also applied in the dSPACE to suppress the inverse compensation error and meanwhile improve the control precision. Finally, the experimental results will be provided to demonstrate the effectiveness of the developed control scheme.

7.1 Identification of the Asymmetric Hysteresis Model and the Inverse Compensation

In Sect. 4.1, we developed an asymmetric hysteresis model ASPI model to describe the hysteresis behaviors in the magnetostrictive actuator. To facilitate the identification, the numerical expression of the ASPI model is written as

$$
\begin{aligned}
u(t) &= P[v](t) + \Psi[v](t) + g(v)(t) \\
&= p_0 v(t) + \sum_{j=1}^{n} p_j F_{r_j}[v](t) + \sum_{j=1}^{m} q_j \Psi_{c_j}[v](t) + g(v)(t) \quad (30.73)
\end{aligned}
$$

where p_j denotes the weight of the play operator; $F_{r_j}[v](t)$ is the play operator at the threshold of r_j; n is the number of the play operator used for identification. q_j denotes the weight of the elementary shift operator; $\Psi_{c_j}[v](t)$ is the elementary shift operator at the slope of c_j; m is the number of the elementary shift operator used for identification. $g(v)(t)$ is an auxiliary function, which is selected as

$$
g(v)(t) = -a_3 v(t)^3 - a_2 v(t)^2 - a_1 v(t) - a_0 \quad (30.74)
$$

The thresholds r_j were selected as $r_j = 0.3j$ $(j = 1, 2, \ldots n)$, where the current is a designed decreasing sine signal as $i(t) = 5e^{-0.1t} \sin(2\pi t)$, see Fig. 30.7a. The weights p_j in (30.73) can be found by the following constrained quadratic optimization:

$$
\min\{[C\Lambda - d]^T [C\Lambda - d]\} \quad (30.75)
$$

with the constraints

$$
p(j) \geq 0, \quad j \in \{0, 1, 2, \ldots, N\} \quad (30.76)
$$

where $\Lambda = [p_0, \ldots p_{m+n+4}]^T$, $C = [F_{r_1}, \ldots, F_{r_N}]$, $N = m + n + 4$, d is the output of the magnetostrictive actuator under a designed amplitude decreasing sinusoidal

a

b

c
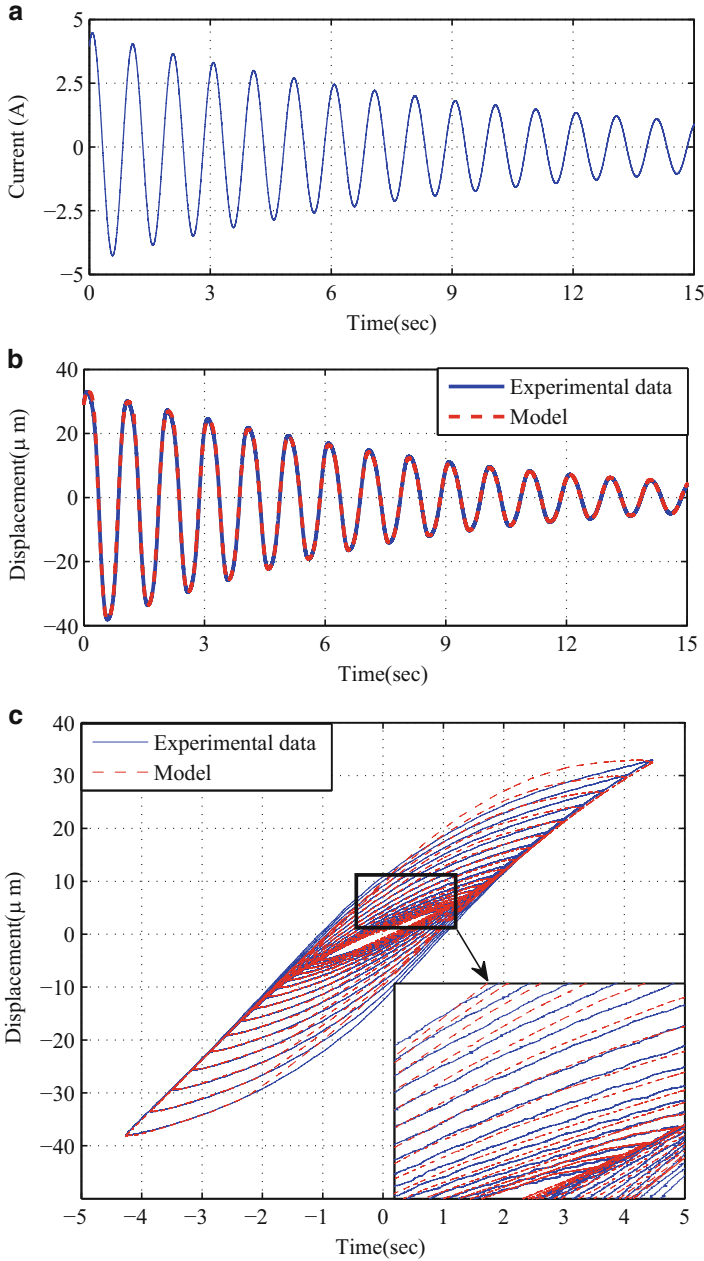
Fig. 30.7 Comparison of experimental data and the model. (**a**) Input signal. (**b**) Comparison of the output between the experimental data and the model. (**c**) Comparison of the input and output responses between the experimental data and the model

Table 30.2 Coefficients of the ASPI model

Numbers	r_j	p_j	c_j	q_j	a_j
0	0	0.9002			0
1	0.3	0.8445	1.1	1.3809	0
2	0.6	0.4276	1.2	0	0.3106
3	0.9	1.4821	1.3	0	0.0417
4	1.2	0.6097	1.4	0	
5	1.5	1.3596	1.5	0	
6	1.8	1.2051	1.6	0	
7	2.1	1.0574	1.7	0	
8	2.4	0.2835	1.8	1.0056	
9	2.7	0.1636			

input signal. Then, the nonlinear least-square optimization toolbox in MATLAB was employed to identify the above parameters and the results are shown in Table 30.2. Figure 30.7b, c show the comparison between the experimental data and the model.

Based on above parameters, the inverse multiplicative structure compensator is therefore implemented as

$$v(t) = \hat{\Pi}^{-1}[u](t) = \hat{P}^{-1}[u - \hat{H}[v]](t) \tag{30.77}$$

where

$$\hat{P}[u]^{-1}(t) = \hat{\bar{p}}_0 u(t) + \sum_{i=1}^{n} \hat{\bar{p}}_i F_{\hat{\bar{r}}_i}[v](t) \tag{30.78}$$

$$\hat{\bar{r}}_i = p_0 \hat{r}_i + \sum_{l=1}^{i} \sum_{j=1}^{l-1} b_j (\hat{r}_l - \hat{r}_{l-1}) \tag{30.79}$$

$$\hat{\bar{p}}_0 = 1/\hat{p}_0 \tag{30.80}$$

$$\hat{\bar{p}}_i = -\frac{\hat{p}_i}{(\hat{p}_0 + \sum_{j=1}^{i} \hat{p}_j)(\hat{p}_0 + \sum_{j=1}^{i-1} \hat{p}_j)} \tag{30.81}$$

Therefore, the thresholds and weights of $P^{-1}[\cdot](t)$ are calculated as: $\hat{\bar{r}}_i[0, 0.2701, 0.7935, 1.4452, 2.5415, 3.82075.5078, 7.5565, 9.9223, 12.3732]$, $\hat{\bar{p}}_0 = 1.1109$, $\hat{\bar{p}}_i = [-0.5377, -0.1128, -0.1867, -0.0391, -0.0567, -0.0314, -0.0196, -0.0044, -0.0024]$. The inverse compensator was implemented in the Matlab/Simulink and the codes were transformed into real-time control codes and downloaded to the dSPACE board. A desired tracking signal $u_d(t) = B_1 \sin(2\pi t)$, $B_1 = 5, 10, 15$ was applied to the compensator. Figures 30.8, 30.9, and 30.10 show the compensation results. It can be seen that, due to the existence of the modeling errors, the

Fig. 30.8 The inverse compensator with desired input $u_d(t) = 5\sin(2\pi t)$
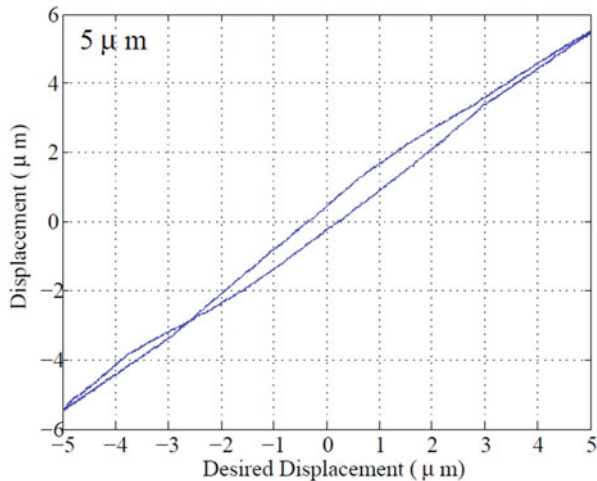
Fig. 30.9 The inverse compensator with desired input $u_d(t) = 10\sin(2\pi t)$
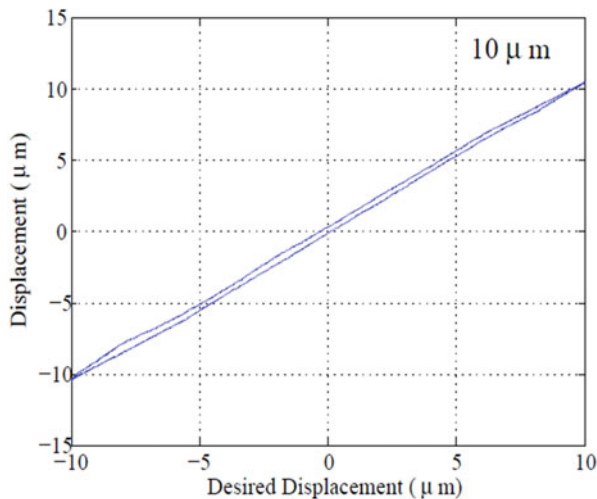
inverse compensation errors are unavoidable, leading to approximate linear relations between the input current and the output displacement. To accommodate compensation errors, the prescribed adaptive control scheme will be applied.

7.2   The Implementation of the Prescribed Adaptive Control

The entire control scheme is illustrated in Fig. 30.4. Since in experiments we only focus on the low frequency application, the dynamic model of the magnetostrictive actuated system is first reduced to a first-order system [30], namely we select

Fig. 30.10 The inverse compensator with desired input $u_d(t) = 15\sin(2\pi t)$

$n = 1$, $Y = x(t)$ in (30.52). Similar treatment can also be found in [31, 32]. The control objective is to force the output of the magnetostrictive-actuated system to follow the desired signal $x_d = 5\sin(t)$ and ensure the transient and steady-state performance of the tracking error within the prescribed function area. The prescribed performance function is selected as $\rho = (1 - 0.07)e^{-t} + 0.07$ with $\overline{M} = 10$, $\underline{M} = -10$. The parameters in the control and adaptive laws are selected as $c_1 = 30$, $\eta_\zeta = 2$, $\Gamma_a = 2$, $\eta_D = 25$. The initial state is chosen as $x(0) = 1.3$. In addition, in the implementation the function $\text{sgn}(z_n)$ is replaced by $\text{sat}(z_n)$ to avoid the chattering effect. The experimental results are shown in Figs. 30.11– 30.14, and 30.15. Figure 30.11 shows the control signal. Figure 30.12 shows the tracking error. It can be seen that a fairly satisfactory tracking performance is achieved and the tracking error converges to a small neighborhood of zero. The input–output relation of the magnetostrictive actuator with prescribed adaptive controller is demonstrated in Fig. 30.13. To further illustrate the effectiveness of the adopted controller, a desired signal $x_d = 5\sin(100 \cdot 2\pi t)$ with higher frequency is applied. The prescribed performance function is selected as $\rho = (1 - 0.07)e^{-100t} + 0.07$ with $\overline{M} = 10$, $\underline{M} = -10$. Figures 30.14 and 30.15 show the tracking error and the input–output relation of the magnetostrictive-actuated dynamic system with the prescribed adaptive controller. From above experimental results, it can be seen that the developed prescribed adaptive controller shows an excellent tracking performance.
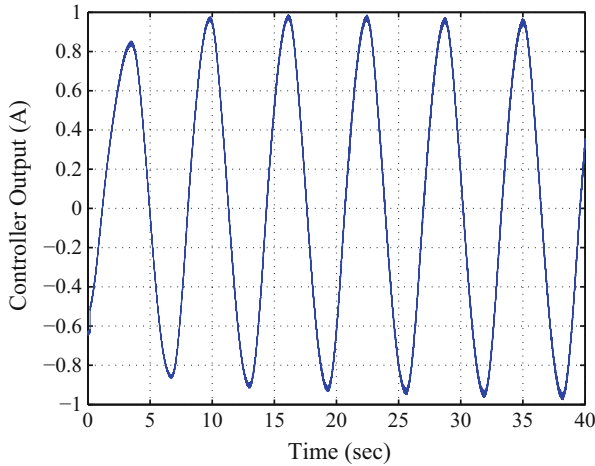
Fig. 30.11 The control input signal to the system
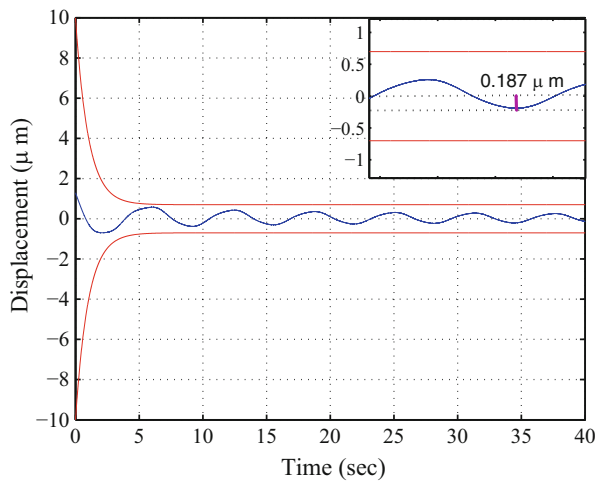
Fig. 30.12 The tracking error

8   Conclusions

This chapter systematically demonstrates an inverse adaptive controller design method for the magnetostrictive-actuated dynamic system. The developed ASPI model is used to describe the asymmetric hysteresis effect, the inverse multiplicative structure is utilized to obtain the inverse compensator of the ASPI model. Due to the presence of the estimated error, the inverse compensation error is unavoidable. By means of the composition theorem, the analytical inverse compensation error
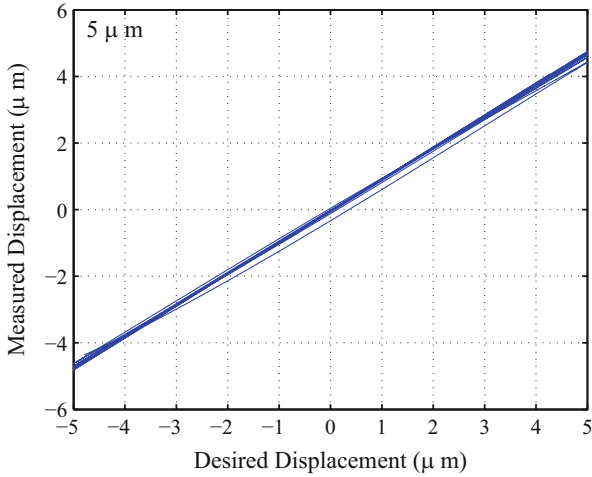
Fig. 30.13 The input–output relation of the magnetostrictive actuator with prescribed adaptive controller
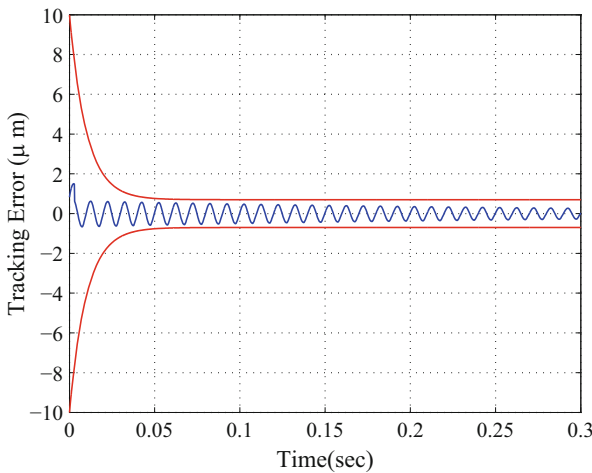
Fig. 30.14 The tracking error with desired input $x_d = 5\sin(100 \cdot 2\pi t)$

is derived. To suppress such a compensation error, a prescribed adaptive control method is applied, in which the global stability of the closed-loop system with a prescribed transient and steady-state performance of the tracking error can be guaranteed. The effectiveness of the proposed control scheme is validated via the experimental tests.
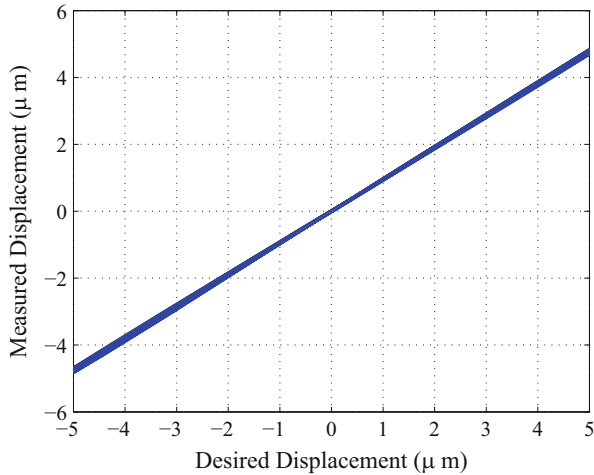
Fig. 30.15 The input–output relation of the magnetostrictive actuator with prescribed adaptive controller under the desired input $x_d = 5\sin(100 \cdot 2\pi t)$

References

1. Davino D, Giustiniani A, Visone C (2011) A two-port nonlinear model for magnetoelastic energy-harvesting devices. IEEE Trans Ind Electron 58(6):2556–2564
2. Olabi AG, Grunwald A (2008) Design and application of magnetostrictive materials. Mater Des 29(2):469–483
3. Yang B-T, Yang D-H, Xu P-Y (2012) Large stroke and nanometer-resolution giant magnetostrictive assembled actuator for driving segmented mirrors in very large astronomical telescopes. Sensor Actuat A-Phys 179(0):193–203
4. Karunanidhi S, Singaperumal M (2010) Design, analysis and simulation of magnetostrictive actuator and its application to high dynamic servo valve. Sensor Actuat A-Phys 157(2):185–197
5. Wang C-L, Cheng X, An P (2010) Modeling and simulation of a high-frequency micro-pump based on giant magnetostrictive material (GMM). J Coal Sci Eng (China) 16(2):206–209
6. Tan XB, Baras JS (2004) Modeling and control of hysteresis in magnetostrictive actuators. Automatica 40(9):1469–1480
7. Li L, Zhang C, Yan B, Zhang L, Li X (2011) Research of fast-response giant magnetostrictive actuator for space propulsion system. IEEE Trans Plasma Sci 39(2):744–748
8. Oates WS, Evans PG, Smith RC, Dapino MJ (2009) Experimental implementation of a hybrid nonlinear control design for magnetostrictive actuators. J Dyn Syst Meas Control 131(4):041004
9. Su C-Y, Wang Q, Chen X, Rakheja S (2005) Adaptive variable structure control of a class of nonlinear systems with unknown Prandtl-Ishlinskii hysteresis. IEEE Trans Autom Control 50(12):2069–2074
10. Krejci P, Kuhnen K (2001) Inverse control of systems with hysteresis and creep. IEE Proc Control Theory Appl 148(3):185–192
11. Kuhnen K (2003) Modeling, identification and compensation of complex hysteretic nonlinearities. Eur J Control 9(4):407–418
12. Li Z, Hu Y, Liu Y, Chen T, Yuan P (2012) Adaptive inverse control of non-linear systems with unknown complex hysteretic non-linearities. IET Control Theory A 6(1):1–7

13. Rakotondrabe M (2011) Bouc-Wen modeling and inverse multiplicative structure to compensate hysteresis nonlinearity in piezoelectric actuators. IEEE Trans Autom Sci Eng 8(2):428–431
14. Zhou J, Wen C, Li T (2012) Adaptive output feedback control of uncertain nonlinear systems with hysteresis nonlinearity. IEEE Trans Autom Control 57(10):2627–2633
15. Preisach F (1935) Über die magnetische Nachwirkung. Z Phys 94:277–302
16. Jiang H, Ji HL, Qiu JH, Chen YS (2010) A modified prandtl-ishlinskii model for modeling asymmetric hysteresis of piezoelectric actuators. IEEE Trans Ultrason Ferroelect Freq Control 57(5):1200–1210
17. Jiang CG, Deng MC, Inoue A (2008) A novel modeling of nonlinear plants with hysteresis described by non-symmetric play operator. The 7th world congress on intelligent control and automation, Chongqing, China, pp. 2221–2224
18. Janaideh AM, Mao JQ, Rakheja S, Xie WF, Su C-Y (2008) Generalized Prandtl-Ishlinskii hysteresis model: hysteresis modeling and its inverse for compensation in smart actuators. IEEE Conference on Decision and Control, Cancun, Mexico, pp. 5182–5187
19. Tao G, Kokotovic PV (1993) Adaptive control of systems with backlash. Automatica 29(2):323–335
20. Tao G, Kokotovic PV (1995) Adaptive control of plants with unknown hysteresis. IEEE Trans Autom Control 40(2):200–212
21. Janaideh AM, Su C-Y, Rakheja S (2012) Inverse compensation error of the Prandtl-Ishlinskii model. IEEE conference on decision and control, Maui, HI, USA, pp. 1597–1602
22. Li Z, Su C-Y, Chen X (2014) Modeling and inverse adaptive control of asymmetric hysteresis systems with applications to magnetostrictive actuator. Control Eng Pract 33:148–160
23. Li J, Xu M (2011) Modified Jiles-Atherton-Sablik model for asymmetry in magnetomechanical effect under tensile and compressive stress. J Appl Phys 110(6):063918
24. Brokate M, Sprekels J (1996) Hysteresis and phase transitions. Springer, New York
25. Krejci P (1986) Hysteresis, convexity and dissipation in hyperbolic equation. International series of math science and applications. Gakkotosho, Tokyo
26. Zhou J, Wen C, Zhang Y (2004) Adaptive backstepping control of a class of uncertain nonlinear systems with unknown backlash-like hysteresis. IEEE Trans Autom Control 49(10):1751–1759
27. Bechlioulis CP, Rovithakis GA (2009) Adaptive control with guaranteed transient and steady state tracking error bounds for strict feedback systems. Automatica 45(2):532–538
28. Bechlioulis CP, Rovithakis GA (2008) Robust adaptive control of feedback linearizable MIMO nonlinear systems with prescribed performance. IEEE Trans Autom Control 53(9):2090–2099
29. Kristic M, Kokotovic PV, Kanellakopoulos I (1995) Nonlinear and adaptive control design. Wiley, New York
30. Riccardi L, Naso D, Turchiano B, Janocha H (2013) Adaptive control of positioning systems with hysteresis based on magnetic shape memory alloys. IEEE Trans Control Syst Technol 21(6):2011–2023
31. Shen J-C, Jywe W-Y, Chiang H-K, Shu Y-L (2008) Precision tracking control of a piezoelectric-actuated system. Precis Eng 32(2):71–78
32. Zhong J, Yao B (2008) Adaptive robust precision motion control of a piezoelectric positioning stage.: IEEE Trans Control Syst Technol 16(5):1039–1046

Index