إلى قارئ هذا الكتاب ، تحية طيبة وبعد ...

لقد أصبحنا نعيش في عالم يعج بالأبحاث والكتب والمعلومات، وأصبح العلم معياراً حقيقياً لتفاضل الأمم والدول والمؤسسات والأشخاص على حدٍّ سواء، وقد أمسى بدوره حلاً شبه وحيدٍ لأكثر مشاكل العالم حدة وخطورة، فالبيئة تبحث عن حلول، وصحة الإنسان تبحث عن حلول، والموارد التي تشكل حاجة أساسية للإنسان تبحث عن حلول كذلك، والطاقة والغذاء والماء جميعها تحديات يقف العلم في وجهها الآن ويحاول أن يجد الحلول لها. فأين نحن من هذا العلم ؟ وأين هو منا؟

نسعى في موقع عالم الإلكترون www.4electron.com لأن نوفر بين أيدي كل من حمل على عاتقه مسيرة درب تملؤه التحديات ما نستطيع من أدوات تساعده في هذا الدرب، من مواضيع علمية، ومراجع أجنبية بأحدث إصداراتها، وساحات لتبادل الآراء والأفكار العلمية والمرتبطة بحياتنا الهندسية، وشروحٍ لأهم برمجيات الحاسب التي تتداخل مع تطبيقات الحياة الأكاديمية والعملية، ولكننا نتوقع في نفس الوقت أن نجد بين الطلاب والمهندسين والباحثين من يسعى مثلنا لتحقيق النفع والفائدة للجميع، ويحلم أن يكون عضواً في مجتمعٍ يساهم بتحقيق بيئة خصبة للمواهب والإبداعات والتألق، فهل تحلم بذلك ؟

حاول أن تساهم بفكرة، بومضة من خواطر تفكيرك العلمي، بفائدة رأيتها في إحدى المواضيع العلمية، بجانب مضيء لمحته خلف ثنايا مفهوم هندسي ما. تأكد بأنك ستلتمس الفائدة في كل خطوة تخطوها، وترى غيرك يخطوها معك ...

أخي القارئ، نرجو أن يكون هذا الكتاب مقدمة لمشاركتك في عالمنا العلمي التعاوني، وسيكون موقعكم عالم الإلكترون ww.4electron.com بكل الإمكانيات المتوفرة لديه جاهزاً على الدوام لأن يحقق البيئة والواقع الذي يبحث عنه كل باحث أو طالب في علوم الهندسة، ويسعى فيه للإفادة كل ساعٍ ، فأهلاً وسهلاً بكم .

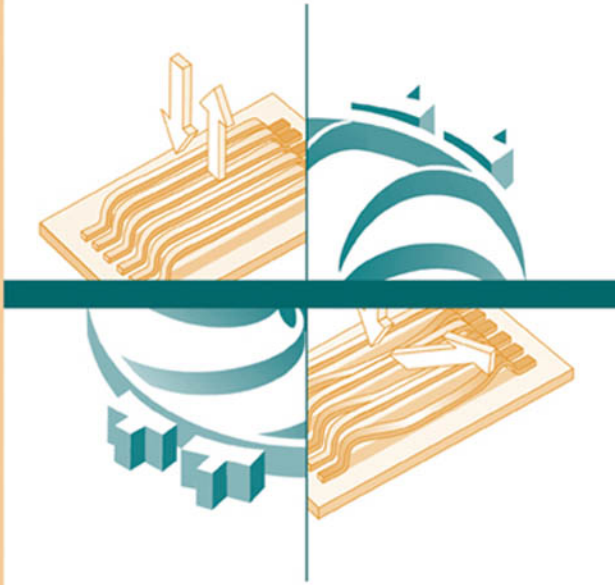
مع تحيات إدارة الموقع وفريق عمله

www.4electron.com

MEMS MICROELECTROMECHANICAL SYSTEMS SERIES

An Introduction to Microelectromechanical Systems Engineering

SECOND EDITION



NADIM MALUF
KIRT WILLIAMS

An Introduction to Microelectromechanical Systems Engineering

Second Edition

For a listing of recent titles in the *Artech House Microelectromechanical Systems* (*MEMS*) *Series*, turn to the back of this book.

An Introduction to Microelectromechanical Systems Engineering

Second Edition

Nadim Maluf
Kirt Williams

Artech House, Inc.
Boston • London
www.artechhouse.com

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data
Maluf, Nadim.
An Introduction to microelectromechanical systems engineering–2nd ed. –(Artech House microelectromechanical library)
1. Microelectromechanical systems
I. Title II. Williams, Kirt
621.3'81

ISBN 1-58053-590-9

Cover design by Igor Valdman

685 Canton Street
Norwood, MA 02062

International Standard Book Number: 1-58053-590-9

10 9 8 7 6 5 4 3 2 1

To our families
Tanya, Ella, and Jad
Erika, Gordon, Brynn, and Reed

Contents

Foreword

According to my best recollection, the acronym for microelectromechanical systems (MEMS) was officially adopted by a group of about 80 zealots at a crowded meeting in Salt Lake City in 1989 called the Micro Tele-Operated Robotics Workshop. I was there to present an invited paper that claimed MEMS should be used to fabricate resonant structures for the purposes of timekeeping, and I was privileged to be part of this group of visionaries for one and a half exciting days. The proceedings may not be in print any longer. However, I recall that they were given an Institute of Electrical and Electronic Engineers (IEEE) catalog number of 89TH0249-3. Discussion at the workshop about the name of this new field of research raged for over an hour, and several acronyms were offered, debated, and rejected. When the dust settled, I recall that Professor Roger Howe of the University of California at Berkeley stood up and announced, "Well, then, the name is MEMS." In this way, the group came to consensus. The research they conducted, unique to any currently being conducted in the United States (or the world for that matter) would hereafter be known as "MEMS."

In those early, heady, exciting, and terribly uncertain days, many issues faced those in the nascent field that researchers today would find hard to remember. For example, our hearty band constantly worried if any scholarly journal would publish the papers we wrote. Sources of research funding were hard to find and difficult to maintain. MEMS fabrication was itself a major issue. Topics of conversation were frequently about the nature, properties, and standardization of the polysilicon that the pioneering band of researchers was using to demonstrate the early, elementary structures of the day. Even the most daring and idealistic of students occasionally turned down the offer to work with the faculty of that era: the work sometimes appeared too farfetched for the taste of even the green-eyed zealots among the graduate student population.

In the 10 years since the momentous events of that watershed workshop, the National Science Foundation (NSF) funded a set of MEMS projects under its "Emerging Technologies Initiative," headed at the time by George Hazelrigg. NSF funding continues to this day. The Defense Advanced Projects Research Agency (DARPA) put nearly $200 million into MEMS research. Numerous MEMS journals have sprung up, and the rate of filing of MEMS patents has reached over 160 per calendar year in 1997. The skeptics that predicted the collapse of the field in 1990 are now confronted with the fact that, in 1997, 80 U.S. were companies in the MEMS field. The combined total world market of MEMS reached approximately $2 billion as well. In addition, the most conservative market studies predict a world MEMS market in excess of $8 billion in 2003. In a phrase, MEMS has arrived.

Despite all the rosy news, there remain significant challenges facing the MEMS field. One of these I call the challenge of the "500 MEMS Companies" and the other, the "10,000 MEMS Designers." For the field to fully take root and become ubiquitous, there must be an unprecedented training of tens of thousands of MEMS engineers. Already, the demand for MEMS experts has far outstripped the ability of academia to train them. The only hope is for existing engineers to learn the basics of MEMS and then go up the MEMS learning curve in the traditional way (i.e., learning by doing).

Here is where this book plays an important, essential role on the national stage. Dr. Nadim Maluf has put together one of the finest MEMS primers that you may find on the bookshelf today. Written in a no-nonsense, clear style, the book brings the practicing engineer and student alike to an understanding of how MEMS are designed and fabricated. Dr. Maluf's book concentrates mostly on how to design and manufacture MEMS. This is to be expected of Dr. Maluf, who has impeccable MEMS credentials. Trained in MEMS for his Ph.D. at Stanford University, Dr. Maluf has spent his postdoctoral career as a practicing MEMS engineer and manager at Lucas NovaSensor, one of the early MEMS companies in the field. His industrial career has been focused both on bringing MEMS products successfully to market and on defending his company's market share against encroachment by other technologies. Because this book is written from Dr. Maluf's practical perspective, this volume is sure to have lasting value to the myriad of engineers and executives who are struggling to find a way into the field of MEMS. This book also will serve as a useful resource for those already in the field who wish to broaden their expertise in MEMS fabrication. When I reviewed the manuscript, I was ready to offer Dr. Maluf a great deal of suggestions and corrections. I was quite humbled to realize that, instead, I was eager to have a copy of the new book on my own shelf. It will serve as a reference for not only myself, but also the students and engineers who frequently ask me, "What book should I buy to learn how to make MEMS?"

Albert ("Al") P. Pisano, Ph.D.
MEMS Program Manager
DARPA

Preface

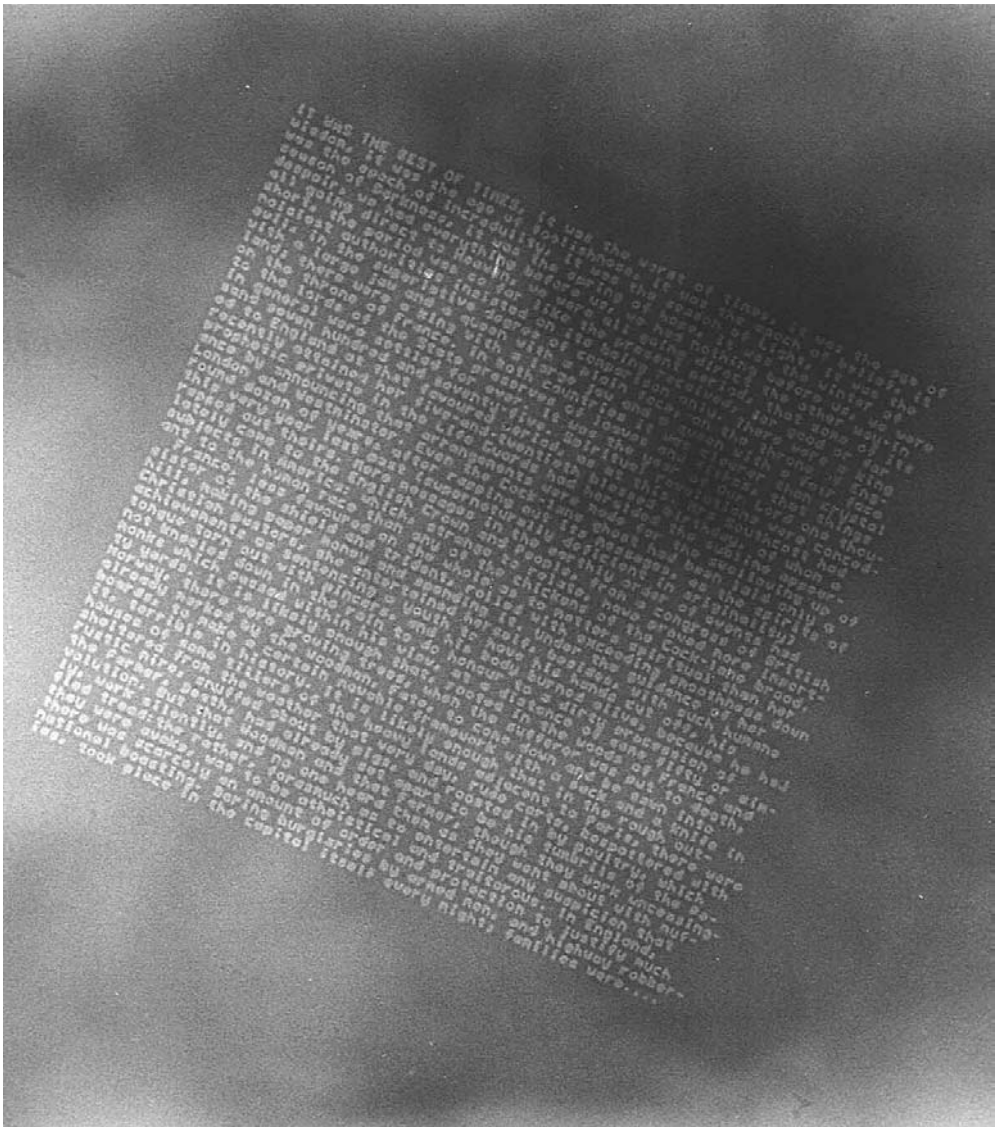
The past few years have witnessed an increasing maturity of the MEMS industry and a rapid introduction of new products addressing applications ranging from biochemical analysis to fiber-optic telecommunications. The market size for MEMS products has doubled in the past 5 years and is projected to grow at this fast rate for the foreseeable future. The corresponding technology has enjoyed a fast pace of development and has rapidly spread to institutions and companies on all inhabited continents. A search of the keyword MEMS in all granted patents in the United States since 1998 returns nearly 4,000 patents and references. Many devices have left universities to go into commercial development, and several have reached the stage of becoming products. It is therefore appropriate to extensively revise the text to incorporate advances in the field, new products, as well as suggestions from the readers.

As we revised the original text and added substantial new material, we strived to retain the style characteristic of an introductory book intended for a broad audience of scientists, engineers, students, and business executives. This revised edition continues to assume that the reader has no prior experience in MEMS technology but does possess an understanding of basic scientific concepts equivalent to first-year college physics and chemistry. The objective remained to introduce a select number of representative demonstrators that are now or are soon to be commercially available. We added many more illustrations and pictures to aid the reader in developing a familiarity with the technology. We also included throughout the text more practical tidbits that are useful to those who wish to apply this technology to their needs.

In this revision, we have expanded on the fabrication processes, adding new methods and materials. The advantages and limitations of many micromachined structures are covered in more detail. We divided the chapter on commercial structures into four chapters, each focusing on a specific application, and then expanded each chapter with appropriate material covering new technical developments and products. Chapter 4 is now specific to automotive and industrial applications, covering traditional products, such as pressure sensors, accelerometers, and yaw-rate sensors, and new emerging products in valving and pumping. Chapter 5 now covers the applications of MEMS in photonics, including displays, optical sensors, and new products that are now common in fiber-optic telecommunications. The focus of Chapter 6 is on applications in life sciences, with emphasis on new products and developments specific to biochemical analysis and microfluidics. With the emergence of wireless and radio frequency (RF) as a new market for MEMS technology, we dedicated Chapter 7 to describe recent developments and introductions in this promising area. In Chapters 4 through 7, we expanded where appropriate on the

application and on the system that includes the MEMS product. We also expanded the material in Chapter 8 on packaging to include packaging of optical MEMS products and added an entirely new section on reliability and quality assurance. We added several references to each chapter to direct the advanced reader to the source of the material. We also expanded the glossary to assist the novice in understanding and relating to a new terminology.

Many people provided us technical information and materials specifically for the second edition of this book. We thank Bardia Pezeshki of Santur Corporation; John (Hal) Jerman of Iolon; Asif Godil of Lightconnect; Greg Ortiz of Surface Technology Systems; Bonnie Gray; Greg Jepson of Bullen Ultrasonics; Chris Bang and Den Feinberg of Microfabrica; Malcom Gower of Exitech; Amy Wang; Brian Paegel of The Scripps Research Institute; Carol Schembri and John Larson of Agilent Technologies; Didier Lacroix and Ken Cioffi of Discera; Michael Cohn of MicroAssembly; Nelson Fuller of Alumina Micro; and Stephen Durant and Christopher Eide of Morrison and Foerster. Evan Green and Carter Hand of New Focus were kind enough to review portions of the manuscript. Thanks go to our editor, Mark Walsh, for his unwavering support. Kirt Williams further thanks his former graduate advisor, Professor Richard S. Muller, for having such a profound effect on his life for introducing him to MEMS.



(*Courtesy of: Engineering & Science,* California Institute of Technology, Pasadena, California.)

"It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness…" from *A Tale of Two Cities* by Charles Dickens, engraved on a thin silicon nitride membrane. The entire page measures a mere 5.9 μm on a side, sufficiently small that 60,000 pages—equivalent to the *Encyclopedia Britannica*—can fit on a pinhead. The work, by T. Newman and R. F. W. Pease of Stanford University, won the Feynman challenge in 1985.

Preface to First Edition

I stood a few years ago before an audience at a customer's facility explaining the merits of micromachining technology. The small conference room was packed, and all ears were attentive. Everyone was eager to learn about this mysterious buzzword, "MEMS." Many in the audience were nodding in a sign of comprehension, but the gazed looks on many faces betrayed them. This experience is not unique; rather, it repeats itself frequently in auditoriums around the world. The technology is simply too broad to be explained in a short lecture. Many technical managers, engineers, scientists, and even engineering students with little or no prior experience in microelectromechanical systems are showing a keen interest in learning about this emerging technology. This book is written for these individuals.

I sought in this book to introduce the technology by describing basic fabrication processes and select examples of devices and microsystems that are either commercially available or show great promise in becoming products in the near future—practical examples from the "real world." The objective is to provide a set of representative cases that can give the reader a global understanding of the technology's foundations and a sense of its diversity. The text describes the basic operation and fabrication of many devices, along with packaging requirements. Inspired by the adage "a picture is worth a thousand words," I have included numerous descriptive schematic illustrations. It is my hope that scanning these illustrations will aid the reader in quickly developing a basic familiarity with the technology. Suggestions at the end of each chapter for further reading and an extensive glossary should supplement the main text.

The following paragraphs present an overview of each chapter in the book.

Chapter 1—MEMS: A Technology from Lilliput. This introductory chapter defines the scope of the technology and the applications it addresses. A short analysis of existing markets and future opportunities is also included.

Chapter 2—The Sandbox: Materials for MEMS. This chapter reviews the properties of materials common in micromachining. The emphasis is on silicon and materials that can be readily deposited as thin films on silicon substrates. Three physical effects—piezoresistivity, piezoelectricity, and thermoelectricity—are described in some detail.

Chapter 3—The Toolbox: Processes for Micromachining. Various fabrication techniques used in semiconductor manufacturing and micromachining are introduced. These include a number of deposition and etch methods, as well as lithography. The discussion on etch methods covers the topics of anisotropic etching, dependence on crystallographic planes, and deep reactive ion etching. Three complete manufacturing process flows are described at the end.

Chapter 4—The Gearbox: Commercial MEM Structures and Systems. This chapter includes descriptions of a select list of commercially available micromachined sensors and actuators. The discussion includes the basic principle of operation and a corresponding fabrication process for each device. Among the devices are pressure and inertial sensors, a microphone, a gas sensor, valves, an infrared imager, and a projection display system.

Chapter 5—The New Gearbox: A Peek into the Future. The discussion in this chapter centers on devices and systems still under development but with significant potential for the future. These include biochemical and genetic analysis systems, high-frequency components, display elements, pumps, and optical switches.

Chapter 6—The Box: Packaging for MEMS. The diverse packaging requirements for MEMS are reviewed in this chapter. The basic techniques of packaging sensors and actuators are also introduced. A few nonproprietary packaging solutions are described.

The writing of a book usually relies on the support and encouragement of colleagues, friends, and family members. This book is no exception. I am grateful to Al Pisano for his general support and for recognizing the value of an introductory book on MEMS. I would like to thank Greg Kovacs, Kirt Williams, and Denise Salles for reading the manuscript and providing valuable feedback. They left an indelible mark of friendship on the pages of the book. I am thankful to many others for their comments, words of encouragement, and contributions. To Bert van Drieënhuizen, Dominik Jaeggi, Bonnie Gray, Jitendra Mohan, John Pendergrass, Dale Gee, Tony Flannery, Dave Borkholder, Sandy Plewa, Andy McQuarrie, Luis Mejia, Stefani Yee, Viki Williams, and the staff at NovaSensor, I say, "Thank you!" Jerry Gist's artistic talents proved important in designing the book cover. For those I inadvertently forgot to mention, please forgive me. I am also grateful to DARPA for providing partial funding under contract N66001-96-C-8631. Last, but not least, words cannot duly express my gratitude and love to my wife, Tanya. She taught me over the course of writing this book the true meaning of love, patience, dedication, understanding, and support. I set out in this book to teach technology, but I finished learning from her about life.

Nadim Maluf

August 1999

CHAPTER 1

MEMS: A Technology from Lilliput

"...And I think to myself, what a wonderful world...oh yeah!"
—*Louis Armstrong*

The Promise of Technology

The ambulance sped down the Denver highway carrying Mr. Rosnes Avon to the hospital. The flashing lights illuminated the darkness of the night, and the siren alerted those drivers who braved the icy cold weather. Mrs. Avon's voice was clearly shaken as she placed the emergency telephone call a few minutes earlier. Her husband was complaining of severe palpitations in his heart and shortness of breath. She sat next to him in the rear of the ambulance and held his hand in silence, but her eyes could not hide her concern and fear. The attending paramedic clipped onto the patient's left arm a small device from which a flexible cable wire led to a digital display that was showing the irregular cardiac waveform. A warning sign in the upper right-hand corner of the display was flashing next to the low blood-pressure reading. In a completely mechanical manner reflecting years of experience, the paramedic removed an adhesive patch from a plastic bag and attached it to Mr. Avon's right arm. The label on the discarded plastic package read "sterile microneedles." Then, with her right hand, the paramedic inserted into the patch a narrow plastic tube, while the fingers of her left hand proceeded to magically play the soft keys on the horizontal face of an electronic instrument. She dialed in an appropriate dosage of a new drug called Nocilis™. Within minutes, the display was showing a recovering cardiac waveform, and the blood pressure warning faded in the dark green color of the screen. The paramedic looked with a smile at Mrs. Avon, who acknowledged with a deep sigh of relief.

Lying in his hospital bed the next morning, Mr. Avon was slowly recovering from the disturbing events of the prior night. He knew that his youthful days were behind him, but the news from his physician that he needed a pacemaker could only cause him anguish. With an electronic stylus in his hand, he continued to record his thoughts and feelings on what appeared to be a synthetic white pad. The pen recognized the pattern of his handwriting and translated it to text for the laptop computer resting on the desk by the window. He drew a sketch of the pacemaker that Dr. Harte showed him in the morning; the computer stored an image of his lifesaving instrument. A little device barely the size of a silver dollar would forever remain in his chest and take control of his heart's rhythm. But a faint smile crossed Mr. Avon's lips when he remembered the doctor mentioning that the pacemaker would monitor his level of physical activity and correspondingly adjust his heart rate. After all, he

might be able to play tennis again. With his remote control, he turned on the projection screen television and slowly drifted back into light sleep.

This short fictional story illustrates how technology can touch our daily lives in so many different ways. The role of miniature devices and systems is not immediately apparent here because they are embedded deep within the application they enable. The circumstances of this story call for such devices on many separate occasions. The miniature yaw-rate sensor in the vehicle stability system ensured that the ambulance did not skid on the icy highway. In the event of an accident, the crash acceleration sensor guaranteed the airbags would deploy just in time to protect the passengers. The silicon manifold absolute pressure (MAP) sensor in the engine compartment helped the engine electronic control unit maintain at the location's high altitude the proper proportions in the mixture of air and fuel. As the vehicle was safely traveling, equally advanced technology in the rear of the ambulance saved Mr. Avon's life. The modern blood pressure sensor clipped onto his arm allowed the paramedic to monitor blood pressure and cardiac output. The microneedles in the adhesive patch ensured the immediate delivery of medication to the minute blood vessels under the skin, while a miniature electronic valve guaranteed the exact dosage. The next day, as the patient lay in his bed writing his thoughts in his diary, the microaccelerometer in the electronic quill recognized the motion of his hand and translated his handwriting into text. Another small accelerometer embedded in his pacemaker would enable him to play tennis again. He also could write and draw at will because the storage capacity of his disk drive was enormous, thanks to miniature read and write heads. And finally, as the patient went to sleep, an array of micromirrors projected a pleasant high-definition television image onto a suspended screen.

Many of the miniature devices listed in this story, in particular the pressure, acceleration, and yaw-rate microsensors and the micromirror display, already exist as commercial products. Ongoing efforts at many companies and laboratories throughout the world promise to deliver, in the not-too-distant future, new and sophisticated miniature components and microsystems. It is not surprising, then, that there is widespread belief in the technology's potential to penetrate in the future far-reaching applications and markets.

What Are MEMS—or MST?

In the United States, the technology is known as *microelectromechanical systems* (MEMS); in Europe, it is called *microsystems technology* (MST). A question asking for a more specific definition is certain to generate a broad collection of replies with few common characteristics other than "miniature." But such apparent divergence in the responses merely reflects the diversity of applications this technology enables, rather than a lack of commonality. MEMS is simultaneously a toolbox, a physical product, and a methodology, all in one:

- It is a portfolio of techniques and processes to design and create miniature systems.
- It is a physical product often specialized and unique to a final application—one can seldom buy a generic MEMS product at the neighborhood electronics store.

- "MEMS is a way of making things," reports the Microsystems Technology Office of the United States DARPA [1]. These "things" merge the functions of sensing and actuation with computation and communication to locally control physical parameters at the microscale, yet cause effects at much grander scales.

Although a universal definition is lacking, MEMS products possess a number of distinctive features. They are miniature *embedded* systems involving one or many *micromachined* components or structures. They *enable* higher level functions, though in and of themselves, their utility may be limited—a micromachined pressure sensor in one's hand is useless, but, under the hood, it controls the fuel-air mixture of the car engine. They often *integrate* smaller functions together into one package for greater utility (e.g., merging an acceleration sensor with electronic circuits for self diagnostics). They can also bring *cost benefits* directly through low unit pricing or indirectly by cutting service and maintenance costs.

Although the vast majority of today's MEMS products are better categorized as components or subsystems, the emphasis in MEMS technology should be on the "systems" aspect. True microsystems may still be a few years away, but their development and evolution relies on the success of today's components, especially as these components are integrated together to perform functions ever increasing in complexity. Building microsystems is an evolutionary process; we spent the last 30 years learning how to build micromachined components, and only recently we began learning about their seamless integration into subsystems and ultimately into complete microsystems.

One notable example is the evolution of crash sensors for airbag safety systems. Early sensors were merely mechanical switches. They later evolved into micromechanical sensors that directly measured acceleration. The current generation of devices integrates electronic circuitry alongside a micromechanical sensor to provide self diagnostics and a digital output. It is anticipated that the next generation of devices will also incorporate the entire airbag deployment circuitry that decides whether to inflate the airbag. As the technology matures, the airbag crash sensor may be integrated one day with micromachined yaw-rate and other inertial sensors to form a complete microsystem responsible for passenger safety and vehicle stability.

Examples of future microsystems are not limited to automotive applications (see Table 1.1). Efforts to develop micromachined components for the control of fluids are just beginning to bear fruit. These could one day lead to the integration of micropumps with microvalves and reservoirs to build new miniature drug delivery systems.

What Is Micromachining?

Micromachining is the set of design and fabrication tools that precisely machine and form structures and elements at a scale well below the limits of our human perceptive faculties—the microscale. Micromachining is the underlying foundation of MEMS fabrication; it is the toolbox of MEMS.

Table 1.1 Examples of Present and Future Application Areas for MEMS

Commercial Applications	Invasive and noninvasive biomedical sensors
	Miniature biochemical analytical instruments
	Cardiac management systems (e.g., pacemakers, catheters)
	Drug delivery systems (e.g., insulin, analgesics)
	Neurological disorders (e.g., neurostimulation)
	Engine and propulsion control
	Automotive safety, braking, and suspension systems
	Telecommunication optical fiber components and switches
	Mass data storage systems
	RF and wireless electronics
	Distributed sensors for condition-based maintenance and monitoring structural health
	Distributed control of aerodynamic and hydrodynamic systems
Military Applications	Inertial systems for munitions guidance and personal navigation
	Distributed unattended sensors for asset tracking, and environmental and security surveillance
	Weapons safing, arming, and fusing
	Integrated microoptomechanical components for identify-friend-or-foe systems
	Head- and night-display systems
	Low-power, high-density mass data storage devices
	Embedded sensors and actuators for condition-based maintenance
	Integrated fluidic systems for miniature propellant and combustion control
	Miniature fluidic systems for early detection of threats from biological and chemical agents
	Electromechanical signal processing for small and low-power wireless communication
	Active, conformable surfaces for distributed aerodynamic control of aircraft

Arguably, the birth of the first micromachined components dates back many decades, but it was the well-established integrated circuit industry that indirectly played an indispensable role in fostering an environment suitable for the development and growth of micromachining technologies. As the following chapters will show, many tools used in the design and manufacturing of MEMS products are "borrowed" from the integrated circuit industry. It should not then be surprising that micromachining relies on silicon as a primary material, even though the technology has certainly been demonstrated using other materials.

Applications and Markets

Present markets are primarily in pressure and inertial sensors, inkjet print heads dominated by the Hewlett-Packard Co. of Palo Alto, California, and high-resolution digital displays with Texas Instruments of Dallas, Texas, being a leader in this market. Future and emerging applications include tire pressure sensing, RF and wireless electronics, fiber optical components, and fluid management and processing devices for chemical microanalysis, medical diagnostics, and drug delivery (see Table 1.1). While estimates for MEMS markets vary considerably, they all show significant present and future growth, reaching aggregate volumes in the many billions of

dollars by 2010 [2–7]. The expected growth stems from technical innovations and acceptance of the technology by an increasing number of end users and customers. A rapid adoption rate of microfluidics, RF, and optical MEMS will cause these applications to grow at a faster pace than the more traditional pressure and acceleration sensing products (see Table 1.2). As a result, the percentage of revenues from automotive applications, which consume large volumes of pressure and acceleration sensors, are projected to decrease even though the automotive market will grow to $1.5 billion in 2007. In 1997, automotive applications accounted for 35% of the total $1.2-billion MEMS market [4], dropping to 26% in 2002 and to 18% in 2007 [6]. It is clear, however, from the data that, because of the lack of a single dominant application—the *killer app*—and the diverse technical requirements of end users, there is not a single MEMS market; rather, there are a collection of markets, many of which are considered niche markets, especially when compared to their semiconductor businesses kin. This fragmentation of the overall market reflects itself onto the large number of small and diverse companies engaged in MEMS. Geographically, the United States and Europe lead the world in the manufacture of MEMS-based products, with Japan trailing (see Table 1.3).

An important action when sizing the market for MEMS is to distinguish between components and systems. For instance, the world market for disposable blood pressure sensors in 2000 was approximately 25 million units totaling $30 million at the

Table 1.2 Analysis and Forecast of Worldwide MEMS Markets (in Millions of U.S. Dollars)

	2002	*2007*
	($ 000,000)	($ 000,000)
Microfluidics	1,404	2,241
Optical MEMS	702	1,826
RF MEMS	39	249
Other actuators	117	415
Inertial sensors	819	1,826
Pressure sensors	546	913
Other sensors	273	830
Total	3,900	8,300

The forecasted compounded annual growth rate (CAGR) between 2002 and 2007 is 16%.
(*Source:* [6]).

Table 1.3 Geographical Distribution of the World MEMS Production Facilities

	Number of Fabs
North America	139
Germany	34
France	20
United Kingdom	14
Benelux	17
Scandinavia	20
Switzerland	14
Rest of Europe	10
Japan	41
Rest of Asia	31

(*Source:* [8].)

component level but about $200 million at the system level. The price differential between the component and the system can readily reach a factor of ten and occasionally higher. Another example is an emerging automotive application for MEMS initiated by the U.S. Congress when it passed the Transportation Recall Enhancement, Accountability and Documentation (TREAD) Act in 2000 requiring warning systems in new vehicles to alert operators when their tires are underinflated (the law was in response to the significant number of fatalities from the Ford/Firestone safety issue). A U.S. federal court directed the National Highway Traffic Safety Administration (NHTSA) in August 2003 to require auto manufacturers to install a direct tire measurement system with a pressure sensor in each wheel [9]. With 16 million new vehicles sold in North America each year, there is suddenly a new market for nearly 70 million pressure sensors totaling approximately $100 million per year. The cost of the total system, which includes electronic circuitry and a wireless link to the dashboard, ranges between $65 to $200 [10], making the market size at the system level well over $1 billion per year.

Forecasting of the MEMS markets has not been without its feckless moments. Poor forecasting of emerging applications has left visible scars on many companies engaged in the development and manufacture of MEMS products. For instance, the worldwide market for airbag crash sensors is estimated today at $150 million, even as these components become standard on all 50 million vehicles manufactured every year around the globe. Market studies conducted in the early 1990s incorrectly estimated the unit asking price of these sensors, neglecting the effect of competition on pricing and artificially inflating the size of the market to $500 million. As a result, many companies rushed to enter the market in the early 1990s only to shutter their programs a few years later.

Marketers also did not fare well in predicting the rapid deflation of the telecommunications economic bubble in 2001 and its Draconian effects on the industry. In the midst of that bubble, studies showed that the markets for optical switches and tunable lasers, two areas that relied heavily on MEMS technology, would soon exceed 10 billions dollars. Venture capitalists poured hundreds of millions of dollars into companies that developed products for fiber-optical telecommunications, many based on various aspects of MEMS technology. Large companies rushed to spend billions in acquiring startup companies with innovative product ideas. With its stock at a historical peak in the year 2000, Nortel Networks of Ontario, Canada, purchased Xros, a startup company in Sunnyvale, California, developing a MEMS-based optical switch fabric, for $3.25 billion in stock. The market for optical switches did not materialize and Nortel ultimately shut down the division. During the same period, JDS Uniphase of San Jose, California, acquired Cronos Integrated Microsystems of Research Triangle Park, North Carolina, a MEMS foundry, for $700 million in stock. JDS Uniphase later divested the division to MEMSCAP of Grenoble, France, for approximately $5 million. Dozens of startup companies met the fate of death as funding dried out and revenues did not grow. But if this doomsday scenario inflicted pain on numerous companies, investors and speculators, it also sowed the seeds of great innovation into the MEMS industry and left a breed of highly competitive and reliable products. The intellectual capital left behind will undoubtedly spur in the near future ideas and products for applications beyond fiber-optical telecommunications.

To MEMS or Not To MEMS?

Like many other emerging technologies with significant future potential, MEMS is subject to a rising level of excitement and publicity. As it evolves and end markets develop, this excessive optimism is gradually moderated with a degree of realism reflecting the technology's strengths and capabilities.

Any end user considering developing a MEMS solution or incorporating one in a design invariably reaches the difficult question of "why MEMS?" The question strikes at the heart of the technology, particularly in view of competing methods (e.g., conventional machining or plastic molding techniques, which do not have recourse to micromachining). For applications that can benefit from existing commercial MEMS products (e.g., pressure or acceleration sensors), the answer to this question relies on the ability to meet required specifications and pricing. But the vast majority of applications require unique solutions that often necessitate the funding and completion of an evaluation or development program. In such situations, a clear-cut answer is seldom easy to establish.

In practice, a MEMS solution becomes attractive if it enables a new function or provides significant cost reduction or both. For instance, medical applications generally seem to focus on added or enabled functionality and improved performance, whereas automotive applications often seek cost reduction. Size reduction can play an important selling role but is seldom sufficient as the sole reason unless it becomes enabling in itself. Naturally, reliability is always a dictated requirement. The decision-making process is further complicated by the fact that MEMS is not a single technology but rather a set of technologies (e.g., surface versus bulk micromachining). At this point, it is beneficial for the end user to become familiar with the capabilities and the limitations of any particular MEMS technology selected for the application in mind. The active participation of the end user allows for the application to drive the technology development rather than the frequent opposite.

Companies seeking MEMS solutions often contract a specialized facility for the design and manufacture of the product. Others choose first to evaluate basic conceptual designs through existing foundry services. A few decide to internally develop a complete design. In the latter case, there is considerable risk that manufacturing considerations are not properly taken into account, resulting in significant challenges in production.

Regardless of how exciting and promising a technology may be, its ultimate realization is invariably dependent on economical success. The end user will justify the technology on the basis of added value, increased productivity, or cost competitiveness, and the manufacturer must show revenues and profits. On both tracks, MEMS technology is able to deliver within a set of realistic expectations that may vary with the end application. A key element to cost competitiveness is *batch fabrication* (that is, the practice of simultaneously manufacturing hundreds or thousands of identical parts, thus diluting the overall impact of fixed costs—including the cost of maintaining expensive cleanroom and assembly facilities) (see Figure 1.1). This is precisely the same approach that has resulted over the last few decades in a dramatic decrease in the price of computer memory chips. Unfortunately, the argument works in reverse too: Small manufacturing volumes will bear the full burden of overhead expenses, regardless of how "enabling" the technology may be.

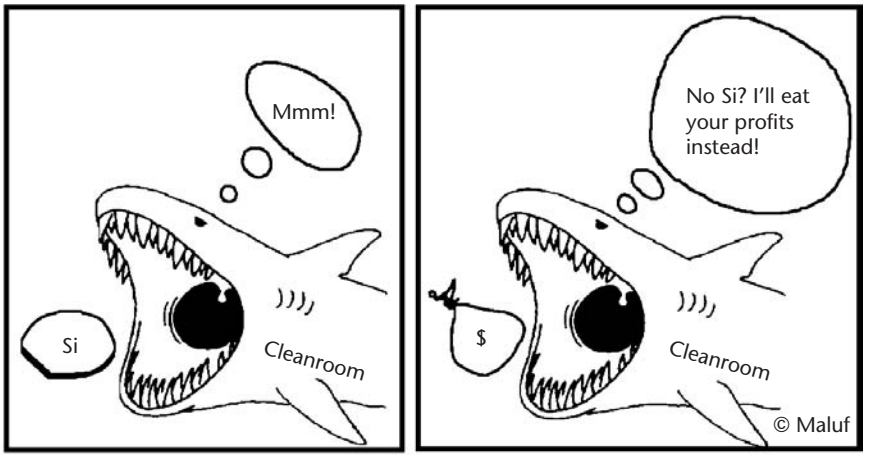


Figure 1.1 Volume manufacturing is essential for maintaining profitability.

Standards

Few disagree that the burgeoning MEMS industry traces many of its roots to the integrated circuit industry. However, the two market dynamics differ greatly with severe implications, one of which is the lack of standards in MEMS. Complementary metal-oxide semiconductor (CMOS) technology has proven itself over the years to be a universally accepted manufacturing process for integrated circuits, driven primarily by the insatiable consumer demand for computers and digital electronics. By contrast, the lack of a dominant MEMS high-volume product (or family of products) and the unique technical requirements of each application have resulted in the emergence of multiple fabrication and assembly processes (the next chapters will introduce them). Standards are generally driven by the needs of high-volume applications, which are few in MEMS. In turn, the lack of standards feeds into the diverging demands of the emerging applications.

The Psychological Barrier

It is human nature to cautiously approach what is new, for it is foreign and untested. Even for the technologically savvy or the fortunate individual living in high-tech regions, there is a need to overcome the comfort zone of the present before engaging the technologies of the future. This cautious behavior translates to slow acceptance of new technologies and derivative products as they get introduced into society. MEMS acceptance is no exception. For example, demonstration of the first micromachined accelerometer took place in 1979 at Stanford University [11]. Yet it took nearly 15 years before it became accepted as a device of choice for automotive airbag safety systems. Naturally, in the process, it was designed and redesigned, tested, and qualified in the laboratory and in the field before it began gaining the confidence of automotive suppliers. The process can be lengthy, especially for embedded systems (see Figure 1.2).

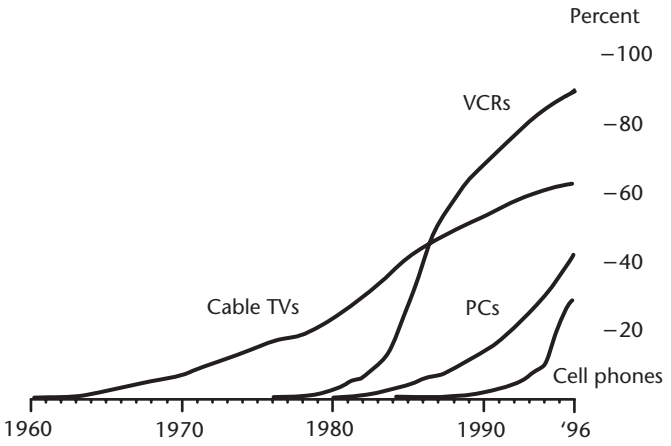


Figure 1.2 The percentage of household penetration of new electronic products. It takes 5 to 15 years before new technologies reach wide acceptance [12].

Today, MEMS and associated product concepts generate plenty of excitement but not without skepticism. Companies exploring for the first time the incorporation of MEMS solutions in their systems do so with trepidation until an internal "MEMS technology champion" emerges to educate the corporation and raise the confidence level. With many micromachined silicon sensors embedded in every car and in numerous critical medical instruments, and with additional MEMS products finding their way into our daily life, the height of this hidden psychological barrier appears to be declining.

Journals, Conferences, and Web Sites

The list of journals and conferences with a focus on micromachining and MEMS continues to grow every year. There is also a growing list of online Web sites, most notably MEMSnet®, an information clearinghouse hosted by the Corporation for National Research Initiatives of Reston, Virginia, and Nexus Association of Grenoble, France, a nonprofit organization with funding from the European Commission. The sites provide convenient links and maintain relevant information directories (see Table 1.4).

List of Journals and Magazines

Several journals and trade magazines published in the United States and Europe cover research and advances in the field. Some examples are:

- *Sensors and Actuators (A, B & C):* a peer-reviewed scientific journal published by Elsevier Science of Amsterdam, The Netherlands.
- *Journal of Micromechanical Systems (JMEMS):* a peer-reviewed scientific journal published by the IEEE of Piscataway, New Jersey, in collaboration with the American Society of Mechanical Engineers (ASME) of New York, New York.

Table 1.4 List of a Few Government and Nongovernment Organizations with Useful On-line Resources

Organization	*Address*	*Description*	*Web Site*
MEMSnet	Reston, VA	U.S. information clearinghouse	www.memsnet.org
MEMS Exchange	Reston, VA	Intermediary broker for foundry services	www.mems-exchange.org
MEMS Industry Group	Pittsburgh, PA	Industrial consortium	www.memsindustrygroup.org
NIST	Gaithersburg, MD	Sponsored U.S. government projects	www.atp.nist.gov
DARPA	Arlington, VA	Sponsored U.S. government projects	www.darpa.mil
IDA	Alexandria, VA	Insertion in military applications	mems.ida.org
NEXUS	Grenoble, France	European MST network	www.nexus-mems.com
VDI/VDE – IT	Teltow, Germany	Association of German Engineers	www.mstonline.de
AIST – MITI	Tokyo, Japan	The "Micromachine Project" in Japan	www.aist.go.jp
ATIP	Albuquerque, NM	Asian Technology Information Project	www.atip.org

- *Journal of Micromechanics and Microengineering (JMM):* a peer-reviewed scientific journal published by the Institute of Physics of Bristol, United Kingdom.
- *Sensors Magazine:* a trade journal with emphasis on practical and commercial applications. It is published by Helmers Publishing, Inc., of Peterborough, New Hampshire.
- *MST News:* a newsletter on microsystems and MEMS. It is published by VDI/VDE Technologiezentrum Informationstechnik GmbH of Teltow, Germany, and is available on-line.
- *Micro/Nano Newsletter:* a publication companion to "*R&D Magazine*" with news and updates on micromachined devices and nanoscale-level technologies. It is published by Reed Business Information of Morris Plains, New Jersey.
- *Small Times Magazine:* a trade journal reporting on MEMS, MST, and nanotechnology. It is published by Small Times Media, LLC, a subsidiary company of Ardesta, LLC, of Ann Arbor, Michigan.

List of Conferences and Meetings

Several conferences cover advances in MEMS or incorporate program sessions on micromachined sensors and actuators. The following list gives a few examples:

- *International Conference on Solid-State Sensors and Actuators (Transducers):* held in odd years and rotates sequentially between North America, Asia, and Europe.
- *Solid-State Sensor and Actuator Workshop (Hilton-Head):* held in even years in Hilton Head Island, South Carolina, and sponsored by the Transducers Research Foundation of Cleveland, Ohio.

- *Micro Electro Mechanical Systems Workshop (MEMS):* an international meeting held annually and sponsored by the IEEE.
- *International Society for Optical Engineering (SPIE):* regular conferences held in the United States and sponsored by SPIE of Bellingham, Washington.
- *Micro Total Analysis Systems (μTAS):* a conference focusing on microanalytical and chemical systems. It is an annual meeting and alternates between North America and Europe.

Summary

Microelectromechanical structures and systems are miniature devices that enable the operation of complex systems. They exist today in many environments, especially automotive, medical, consumer, industrial, and aerospace. Their potential for future penetration into a broad range of applications is real, supported by strong development activities at many companies and institutions. The technology consists of a large portfolio of design and fabrication processes (a toolbox), many borrowed from the integrated circuit industry. The development of MEMS is inherently interdisciplinary, necessitating an understanding of the toolbox as well as of the end application.

References

[1] Dr. Albert Pisano, in presentation material distributed by the U.S. DARPA, available at http://www.darpa.mil.

[2] System Planning Corporation, "Microelectromechanical Systems (MEMS): An SPC Market Study," January 1999, 1429 North Quincy Street, Arlington, VA 22207.

[3] Frost and Sullivan, "World Sensors Market: Strategic Analysis," Report 5509-32, February 1999, 2525 Charleston Road, Mountain View, CA 94043, http://www.frost.com.

[4] Frost and Sullivan, "U.S. Microelectromechanical Systems (MEMS)," Report 5549-16, June 1997, 2525 Charleston Road, Mountain View, CA 94043, http://www.frost.com.

[5] Intechno Consulting, "Sensors Market 2008," Steinenbachgaesslein 49, CH-4051, Basel, Switzerland, http://www.intechnoconsulting.com.

[6] In-Stat/MDR, "Got MEMS? Industry Overview and Forecast," Report IN030601EA, August 2003, 6909 East Greenway Parkway, Suite 250, Scottsdale, AZ 85254, http://www.instat.com.

[7] WTC Wicht Technologie Consulting, "The RF MEMS Market 2002–2007," Frauenplatz 5, D-80331 München, Germany, http://www.wtc-consult.de.

[8] Yole Développement, "World MEMS Fab," 45 Rue Sainte Geneviève, 69006 Lyon, France, http://www.yole.fr.

[9] Public Citizen, Inc., et al. *v.* Norman Mineta, Secretary of Transportation, Docket No. 02-4237, August 6, 2003, United States Court of Appeals, Second Circuit, New York, http://www.ca2.uscourts.gov.

[10] "IC Makers Gear Up for New Tire Pressure Monitor Rule," *Electronic Engineering Times*, December 1, 2003, p. 1.

[11] Roylance, L. M., and J. B. Angell, "A Batch Fabricated Silicon Accelerometer," *IEEE Trans. Electron Devices*, Vol. 26, No. 12, 1979, pp. 1911–1917.

[12] Mercer Management Consulting, Inc., "New Technologies Take Time," *Business Week*, April 19, 1999, p. 8.

Selected Bibliography

Angell, J. B., S. C. Terry, and P. W. Barth, "Silicon Micromechanical Devices," *Scientific American*, Vol. 248, No. 4, April 1983, pp. 44–55.

Gabriel, K. J., "Engineering Microscopic Machines," *Scientific American*, Vol. 273, No. 3, September 1995, pp. 150–153.

Micromechanics and MEMS: Classic and Seminal Papers to 1990, W. S. Trimmer (ed.), New York: Wiley-IEEE Press, 1997.

"Nothing but Light," *Scientific American*, Vol. 279, No. 6, December 1998, pp. 17–20.

Petersen, K. E., "Silicon As a Mechanical Material," *Proceedings of the IEEE*, Vol. 70, No. 5, May 1982, pp. 420–457.

CHAPTER 2

Materials for MEMS

"You can't see it, but it's everywhere you go."
—*Bridget Booher, journalist, on silicon*

If we view micromachining technology as a set of generic tools, then there is no reason to limit its use to one material. Indeed, micromachining has been demonstrated using silicon, glass, ceramics, polymers, and compound semiconductors made of group III and V elements, as well as a variety of metals including titanium and tungsten. Silicon, however, remains the material of choice for microelectromechanical systems. Unquestionably, this popularity arises from the large momentum of the electronic integrated circuit industry and the derived economic benefits, not least of which is the extensive industrial infrastructure. The object of this chapter is to present the properties of silicon and several other materials, while emphasizing that the final choice of materials is determined by the type of application and economics.

Silicon-Compatible Material System

The silicon-compatible material system encompasses, in addition to silicon itself, a host of materials commonly used in the semiconductor integrated circuit industry. Normally deposited as thin films, they include silicon oxides, silicon nitrides, and silicon carbides, metals such as aluminum, titanium, tungsten, and copper, and polymers such as photoresist and polyimide.

Silicon

Silicon is one of very few materials that is economically manufactured in single-crystal substrates. This crystalline nature provides significant electrical and mechanical advantages. The precise modulation of silicon's electrical conductivity using impurity doping lies at the very core of the operation of electronic semiconductor devices. Mechanically, silicon is an elastic and robust material whose characteristics have been very well studied and documented (see Table 2.1). The tremendous wealth of information accumulated on silicon and its compounds over the last few decades has made it possible to innovate and explore new areas of application extending beyond the manufacturing of electronic integrated circuits. It becomes evident that silicon is a suitable material platform on which electronic, mechanical, thermal, optical, and even fluid-flow functions can be integrated. Ultrapure, electronic-grade silicon wafers available for the integrated circuit industry are common today in MEMS. The relatively low cost of these substrates

Table 2.1 Properties of Selected Materials

Property[a]	*Si*	SiO_2	Si_3N_4	*Quartz*	*SiC*	*Diamond*	*GaAs*	*AlN*	*92% Al_2O_3*	*Polyimide*	*PMMA*
Relative permittivity (ε_r)	11.7	3.9	4–8	3.75	9.7	5.7	13.1	8.5	9	—	—
Dielectric strength (V/cm $\times 10^6$)	0.3	5–10	5–10	25–40	4	10	0.35	13	11.6	1.5–3	0.17
Electron mobility (cm^2/V·s)	1,500	—	—	—	1,000	2,200	8,800	—	—	—	—
Hole mobility (cm^2/V·s)	400	—	—	—	40	1,600	400	—	—	—	—
Bandgap (eV)	1.12	8-9	—	—	2.3–3.2	5.5	1.42	—	—	—	—
Young's modulus (GPa)	160	73	323	107	450	1,035	75	340	275	2.5	3
Yield/fracture strength (GPa)	7	8.4	14	9	21	>1.2	3	16	15.4	0.23	0.06
Poisson's ratio	0.22	0.17	0.25	0.16	0.14	0.10		0.31	0.31	0.34	—
Density (g/cm^3)	2.4	2.2	3.1	2.65	3.2	3.5	5.3	3.26	3.62	1.42	1.3
Coefficient of thermal expansion (10^{-6}/°C)	2.6	0.55	2.8	0.55	4.2	1.0	5.9	4.0	6.57	20	70
Thermal conductivity at 300K (W/m·K)	157	1.4	19	1.4	500	990–2,000	0.46	160	36	0.12	0.2
Specific heat (J/g·K)	0.7	1.0	0.7	0.787	0.8	0.6	0.35	0.71	0.8	1.09	1.5
Melting temperature (°C)	1,415	1,700	1,800	1,610	1,800[b]	3,652[b]	1,237	2,470	1,800	380[c]	90[c]

[a]Properties can vary with crystal direction, crystal structure, and grain size.

[b]Sublimates before melting.

[c]Glass transition temperature given for polymers.

(approximately $10 for a 100-mm-diameter wafer and $15 for a 150-mm wafer) makes them attractive for the fabrication of micromechanical components and systems.

Silicon as an element exists with three different microstructures: *crystalline*, *polycrystalline*, or *amorphous*. Polycrystalline, or simply "polysilicon," and amorphous silicon are usually deposited as thin films with typical thicknesses below 5 μm. Crystalline silicon substrates are commercially available as circular wafers with 100-mm (4-in) and 150-mm (6-in) diameters. Larger-diameter (200-mm and 300-mm) wafers, used by the integrated circuit industry, are currently economically unjustified for MEMS. Standard 100-mm wafers are nominally 525 μm thick, and 150-mm wafers are typically 650 μm thick. Double-side-polished wafers commonly used for micromachining on both sides of the wafer are approximately 100 μm thinner than standard thickness substrates.

Visualization of crystallographic planes is key to understanding the dependence of material properties on crystal orientation and the effects of plane-selective etch solutions (see Figure 2.1). Silicon has a diamond-cubic crystal structure that can be

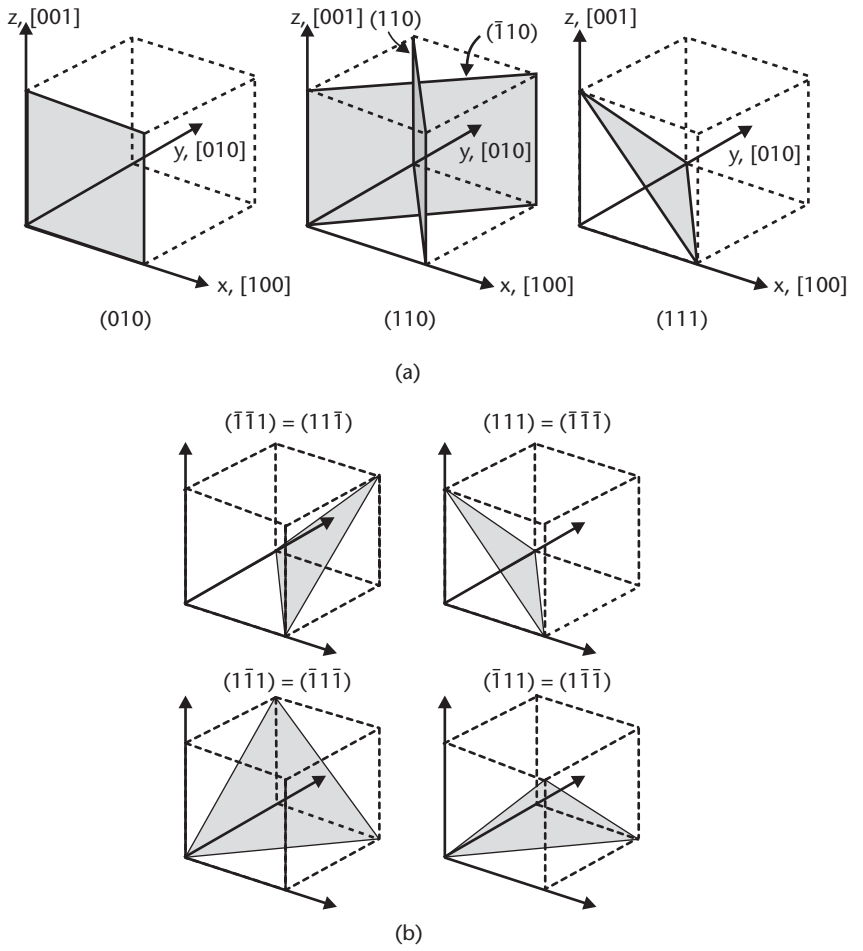


Figure 2.1 (a) Three crystallographic planes and their Miller indices for a simple cubic crystal. Two planes in the {110} set of planes are identified. (b) The four planes in the {111} family. Note that $(1\bar{1}\bar{1})$ is the same plane as (111).

discussed as if it were simple cubic. In other words, the primitive unit—the smallest repeating block—of the crystal lattice resembles a cube. The three major coordinate axes of the cube are called the *principal axes*. Specific directions and planes within the crystal are designated in reference to the principal axes using *Miller indices* [1], a special notation from materials science that, in cubic crystals, includes three integers with different surrounding "punctuation." Directions are specified by brackets; for example [100], which is a vector in the $+x$ direction, referred to the three principal axes (x,y,z) of the cube. No commas are used between the numbers, and negative numbers have a bar over the number rather than a minus sign. Groups of directions with equivalent properties are specified with carets (e.g., <100>, which covers the $[100]=+x, [\bar{1}00]=-x, [010]=+y, [0\bar{1}0]=-y, [001]=+z$, and $[00\bar{1}]=-z$ directions). Parentheses specify a plane that is perpendicular to a direction with the same numbers; for example, (111) is a plane perpendicular to the [111] vector (a diagonal vector through the farthest corner of the unit cube). Braces specify all equivalent planes; for example, $\{11\bar{1}\}$ represents the four equivalent crystallographic planes (111), $(\bar{1}11)$, $(1\bar{1}1)$, and $(11\bar{1})$.

The determinants of plane and direction equivalence are the symmetry operations that carry a crystal lattice (including the primitive unit) back into itself (i.e., the transformed lattice after the symmetry operation is complete is identical to the starting lattice). With some thought, it becomes evident that 90° rotations and mirror operations about the three principal axes are symmetry operations for a simple cubic crystal. Therefore, the $+x$ direction is equivalent to the $+y$ direction under a 90° rotation; the $+y$ direction is equivalent to the $-y$ direction under a mirror operation, and so forth. Hence, the $+x$, $-x$, $+y$, $-y$, $+z$, and $-z$ directions are all equivalent. Vector algebra (using a dot product) shows that the angles between {100} and {110} planes are 45° or 90°, and the angles between {100} and {111} planes are 54.7° or 125.3°. Similarly, {111} and {110} planes can intersect each other at 35.3°, 90°, or 144.7°. The angle between {100} and {111} planes is of particular importance in micromachining because many alkaline aqueous solutions, such as potassium hydroxide (KOH), selectively etch the {100} planes of silicon but not the {111} planes (discussed in detail in Chapter 3). The etch results in cavities that are bounded by {111} planes (see Figure 2.2).

Material manufacturers cut thin circular wafers from large silicon boules along specific crystal planes. The cut plane—the top surface of the wafer—is known as the orientation cut. The (100) wafers dominate in both MEMS and CMOS technology, but wafers are also readily available with (111) orientation and, to a lesser degree, (110) orientation. It should be noted that saying that the surface of a wafer has a particular orientation such as (100) is arbitrary; any orientation within the equivalent {100} group of planes, such as (001), can alternatively be selected. It should be further noted that when referring to the wafer surface (e.g., (100)), the group of planes (e.g., {100}) or direction normal to the surface (e.g., [100]) is often used instead; all are intended to mean the same thing. The (100) and (111) wafers, with *n*- and *p*-type doing, are produced with a minor flat at a specific location relative to a wider, major flat, as shown in Figure 2.2.

Crystalline silicon is a hard and brittle material deforming elastically until it reaches its yield strength, at which point it breaks. Its tensile yield strength is 7 GPa, which is equivalent to a 700-kg (1,500-lb) weight suspended from a 1-mm^2 area. Its Young's modulus is dependent on crystal orientation, being 169 GPa in <110> directions and 130 GPa in <100> directions—near that of steel. The dependence of the mechanical properties on crystal orientation is reflected in the way a silicon wafer preferentially cleaves along crystal planes[1]. While large silicon wafers tend to be fragile, individual dice with dimensions on the order of 1 cm × 1 cm or less are rugged and can sustain relatively harsh handling conditions. As a direct consequence of being a single crystal, mechanical properties are uniform across wafer lots, and wafers are free of intrinsic stresses. This helps to minimize the number of design iterations for silicon transducers that rely on stable mechanical properties for their operation. Bulk mechanical properties of crystalline silicon are largely independent

1. A (100) silicon wafer can be cleaved by scratching the surface with a sharp diamond scribe along a <110> direction (parallel or perpendicular to the flat), clamping the wafer on one side of the scratch, and applying a bending force to the free side of the wafer. Fracture occurs preferentially along <110> directions on the surface. The newly exposed fracture surfaces tend to be {111} planes, which are sloped at 54.7° with respect to the surface.

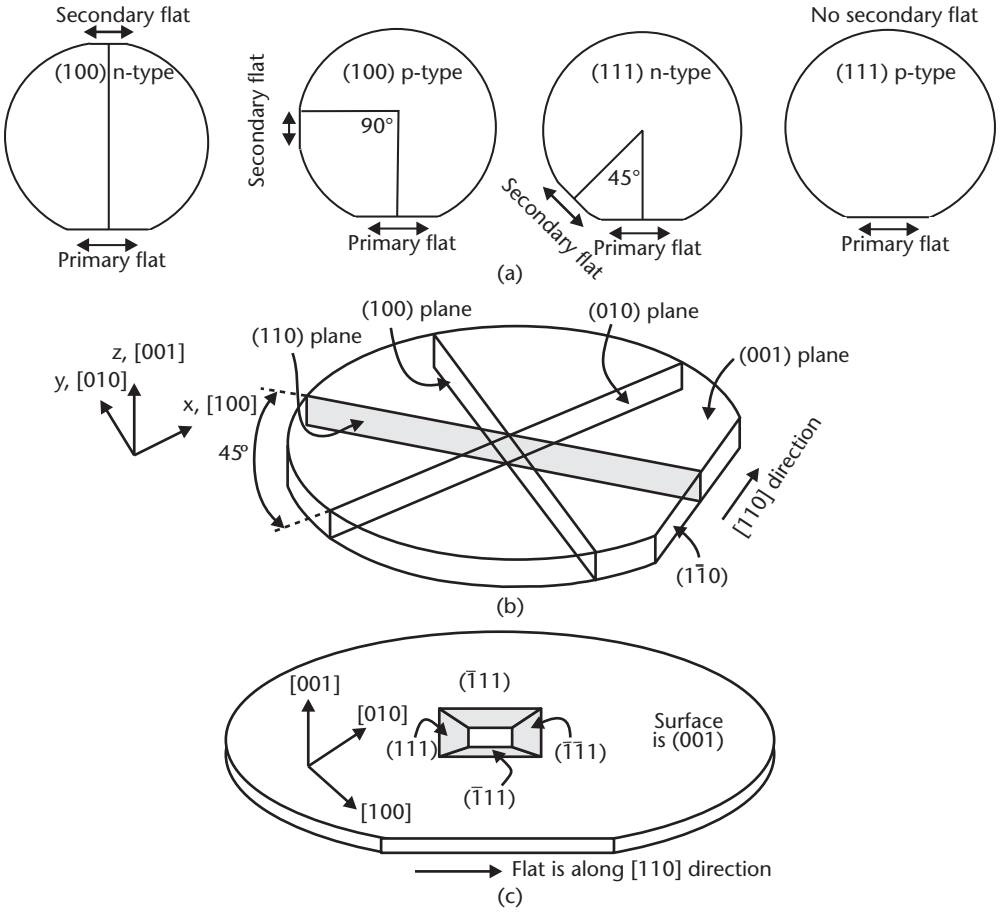


Figure 2.2 (a) Illustration showing the primary and secondary flats of {100} and {111} wafers for both *n*-type and *p*-type doping (SEMI standard); (b) illustration identifying various planes in a wafer of {100} orientation (the wafer thickness is exaggerated); and (c) perspective view of a {100} wafer and a KOH-etched pit bounded by {111} planes.

of impurity doping, but stresses tend to rise when dopant concentrations reach high levels (~ 10^{20} cm^{-3}).

Polysilicon is an important material in the integrated circuit industry and has been extensively studied. A detailed description of its electrical properties is found in [2]. Polysilicon is an equally important and attractive material for MEMS. It has been successfully used to make micromechanical structures and to integrate electrical interconnects, thermocouples, *p-n* junction diodes, and many other electrical devices with micromechanical structures. The most notable example is the acceleration sensor available from Analog Devices, Inc., of Norwood, Massachusetts, for automotive airbag safety systems. Surface micromachining based on polysilicon is today a well-established technology for forming thin (a few micrometers) and planar devices.

The mechanical properties of polycrystalline and amorphous silicon vary with deposition conditions, but, by and large, they are similar to that of single crystal silicon [3]. Both normally have relatively high levels of intrinsic stress (hundreds of MPa) after deposition, which requires annealing at elevated temperatures (>900°C).

Beam structures made of polycrystalline or amorphous silicon that have not been subjected to a careful stress annealing step can curl under the effect of intrinsic stress.

Silicon is a very good thermal conductor with a thermal conductivity greater than that of many metals and approximately 100 times larger than that of glass. In complex integrated systems, the silicon substrate can be used as an efficient heat sink. This feature will be revisited when we review thermal-based sensors and actuators.

Unfortunately, silicon is not an active optical material—silicon-based lasers do not exist. Because of the particular interactions between the crystal atoms and the conduction electrons, silicon is effective only in detecting light; emission of light is very difficult to achieve. At infrared wavelengths above 1.1 μm, silicon is transparent, but at wavelengths shorter than 0.4 μm (in the blue and ultraviolet portions of the spectrum), it reflects over 60% of the incident light (see Figure 2.3). The attenuation depth of light in silicon (the distance light travels before the intensity drops to 36% of its initial value) is 2.7 μm at 633 nm (red) and 0.2 μm at 436 nm (blue-violet). The slight attenuation of red light relative to other colors is what gives thin silicon membranes their translucent reddish tint.

Silicon is also well known to retain its mechanical integrity at temperatures up to about 700°C [4]. At higher temperatures, silicon starts to soften and plastic deformation can occur under load. While the mechanical and thermal properties of polysilicon are similar to those of single crystal silicon, polysilicon experiences slow stress annealing effects at temperatures above 250°C, making its operation at elevated temperatures subject to long-term instabilities, drift, and hysteresis effects. Some properties of silicon at and above room temperature are given in Table 2.2.

The surface of silicon oxidizes immediately upon exposure to the oxygen in air (referred to as *native oxide*). The oxide thickness self-limits at a few nanometers at room temperature. As silicon dioxide is very inert, it acts as a protective layer that prevents chemical reactions with the underlying silicon.

The interactions of silicon with gases, chemicals, biological fluids, and enzymes remain the subject of many research studies, but, for the most part, silicon is considered stable and resistant to many elements and chemicals typical of daily

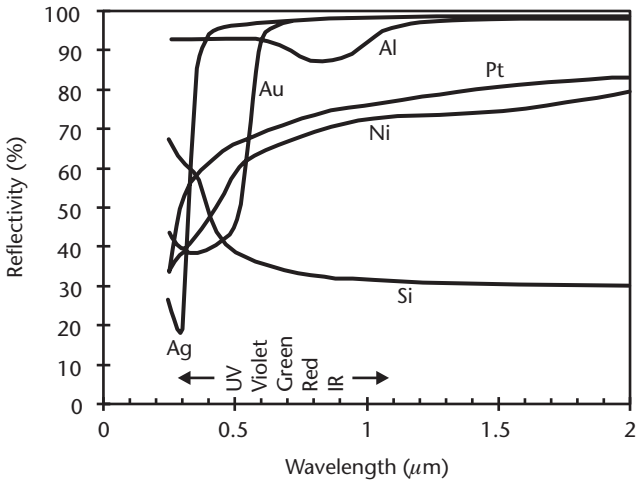


Figure 2.3 Optical reflectivity for silicon and selected metals.

Table 2.2 Temperature Dependence of Some Material Properties of Crystalline Silicon

	300K	*400K*	*500K*	*600K*	*700K*
Coefficient of linear expansion (10^{-6} K^{-1})	2.616	3.253	3.614	3.842	4.016
Specific heat (J/g·K)	0.713	0.785	0.832	0.849	0.866
Thermal conductivity (W/cm·K)	1.56	1.05	0.8	0.64	0.52
Temperature coefficient of Young's modulus (10^{-6} K^{-1})	–90	–90	–90	–90	–90
Temperature coefficient of piezoresistance (10^{-6} K^{-1}) (doping <10^{18} cm^{-3})	–2,500	–2,500	–2,500	—	—
Temperature coefficient of permittivity (10^{-6} K^{-1})	1,000	—	—	—	—

(*Source:* [5].)

applications. For example, experiments have shown that silicon remains intact in the presence of Freon™ gases as well as automotive fluids such as brake fluids. Silicon has also proven to be a suitable material for applications such as valves involving the delivery of ultra-high-purity gases. In medicine and biology, studies are ongoing to evaluate silicon for medical implants. Preliminary medical evidence indicates that silicon is benign in the body and does not release toxic substances when in contact with biological fluids; however, it appears from recent experiments that bare silicon surfaces may not be suitable for high-performance polymerase chain reactions (PCR) intended for the amplification of genetic DNA material.

Silicon Oxide and Nitride

It is often argued that silicon is such a successful material because it has a stable oxide that is electrically insulating—unlike germanium, whose oxide is soluble in water, or gallium arsenide, whose oxide cannot be grown appreciably. Various forms of silicon oxides (SiO_2, SiO_x, silicate glass) are widely used in micromachining due to their excellent electrical and thermal insulating properties. They are also used as sacrificial layers in surface micromachining processes because they can be preferentially etched in hydrofluoric acid (HF) with high selectivity to silicon. Silicon dioxide (SiO_2) is thermally grown by oxidizing silicon at temperatures above 800°C, whereas the other forms of oxides and glass are deposited by chemical vapor deposition, sputtering, or even spin-on (the various deposition methods will be described in the next chapter). Silicon oxides and glass layers are known to soften and flow when subjected to temperatures above 700°C. A drawback of silicon oxides is their relatively large intrinsic stresses, which are difficult to control. This has limited their use as materials for large suspended beams or membranes.

Silicon nitride (Si_xN_y) is also a widely used insulating thin film and is effective as a barrier against mobile ion diffusion—in particular, sodium and potassium ions found in biological environments. Its Young's modulus is higher than that of silicon and its intrinsic stress can be controlled by the specifics of the deposition process. Silicon nitride is an effective masking material in many alkaline etch solutions.

Thin Metal Films

The choice of a thin metal film depends greatly on the nature of the final application. Thin metal films are normally deposited either by sputtering, evaporation, or chemical vapor deposition; gold, nickel, and Permalloy™ (Ni_xFe_y), and a few other metals can also be electroplated. Table 2.3 lists some metals and conducting compounds used as thin films, along with their resistivities (resistivity varies with deposition conditions and is usually higher for thin films than for bulk material).

For basic electrical interconnections, aluminum (usually with a few percent silicon and perhaps copper) is most common and is relatively easy to deposit by sputtering, but its operation is limited to noncorrosive environments and to temperatures below 300°C. For higher temperatures and harsher environments, gold, titanium, and tungsten are substitutes. Aluminum tends to anneal over time and with temperature, causing changes in its intrinsic stresses. As a result, it is typically located away from stress- or strain-sensing elements. Aluminum is a good light reflector in the visible, and gold excels in the infrared. Platinum and palladium are two very stable materials for electrochemistry, though their fabrication entails some added complexity. Gold, platinum, and iridium are good choices for microelectrodes, used in electrochemistry and in sensing biopotentials. Silver is also useful in electrochemistry. Chromium, titanium, and titanium-tungsten are frequently used as very thin (5–20 nm) adhesion layers for metals that have poor adhesion to silicon, silicon dioxide, and silicon nitride. Metal bilayers consisting of an adhesion layer (e.g., chromium) and an

Table 2.3 List of Selected Metals That Can Be Deposited As Thin Films (Up to a Few μm in Thickness) with Corresponding Electrical Resistivities and Typical Areas of Application

Metal	*ρ ($\mu\Omega$·cm)*	*Typical Areas of Application*
Ag	1.58	Electrochemistry
Al	2.7	Electrical interconnects; optical reflection in the visible and the infrared
Au	2.4	High-temperature electrical interconnects; optical reflection in the infrared; electrochemistry; corrosion-resistant contact; wetting layer for soldering
Cr	12.9	Intermediate adhesion layer
Cu	1.7	Low-resistivity electrical interconnects
Indium-tin oxide (ITO)	300–3,000	Transparent conductive layer for liquid crystal displays
Ir	5.1	Electrochemistry; microelectrodes for sensing biopotentials
Ni	6.8	Magnetic transducing; solderable layer
NiCr	200–500	Thin-film laser trimmed resistor; heating element
Pd	10.8	Electrochemistry; solder-wetting layer
Permalloy™ (Ni_xFe_y)	—	Magnetic transducing
Pt	10.6	Electrochemistry; microelectrodes for sensing biopotentials; solderable layer
SiCr	2,000	Thin-film laser trimmed resistor
SnO_2	5,000	Chemoresistance in gas sensors
TaN	300–500	Negative temperature coefficient of resistance (TCR) thin-film laser trimmed resistor
Ti	42	Intermediate adhesion layer
TiNi	80	Shape-memory alloy actuation
TiW	75–200	Intermediate adhesion layer; near zero TCR
W	5.5	High-temperature electrical interconnects; thermionic emitter

intermediate nickel or platinum layer are normally used to solder with silver-tin or tin-lead alloys. For applications requiring transparent electrodes, such as liquid-crystal displays, indium-tin-oxide (ITO) meets the requirements. Finally, Permalloy™ has been explored as a material for thin magnetic cores.

Polymers

Polymers, in the form of polyimides or photoresist, can be deposited with varying thicknesses from a few nanometers to hundreds of microns. Standard photoresist is spin-coated to a thickness of 1 μm to10 μm, but special photoresists such as the epoxy-based SU-8 [6] can form layers up to 100 μm thick. Hardening of the resist under ultraviolet light produces rigid structures. Spin-on organic polymers are generally limited in their application as a permanent part of MEMS devices because they shrink substantially as the solvent evaporates, and because they cannot sustain temperatures above 200°C. Because of their unique absorption and adsorption properties, polymers have gained acceptance in the sensing of chemical gases and humidity [7].

Other Materials and Substrates

Over the years, micromachining methods have been applied to a variety of substrates to fabricate passive microstructures as well as transducers. Fabrication processes for glass and quartz are mature and well established, but for other materials, such as silicon carbide, new techniques are being explored and developed. In the process, these activities add breadth to micromachining technology and enrich the inventory of available tools. The following sections briefly review the use of a few materials other than silicon.

Glass and Fused Quartz Substrates

Glass is without a doubt a companion material to silicon; the two are bonded together figuratively and literally in many ways. Silicon originates from processed and purified silicates (a form of glass), and silicon can be made to bond electrostatically to Pyrex® glass substrates—a process called anodic bonding and common in the making of pressure sensors. But like all relatives, differences remain. Glasses generally have different coefficients of thermal expansion than silicon (fused quartz is lower, while window glass is higher), resulting in interfacial stresses between bonded silicon and glass substrates.

Micromachining of glass and fused quartz (amorphous silicon dioxide) substrates is practical in special applications, such as when an optically transparent or an electrically insulating substrate is required. Crystalline quartz (as opposed to fused quartz) also has the distinct property of being piezoelectric and is used for some MEMS devices. However, micromachining of glass or quartz is limited in scope relative to silicon. Etching in HF or ultrasonic drilling typically yields coarsely defined features with poor edge control. Thin metal films can be readily deposited on glass or quartz substrates and defined using standard lithographic techniques. Channels microfabricated in glass substrates with thin metal microelectrodes have been useful in making capillaries for miniaturized biochemical analysis systems.

Silicon Carbide and Diamond

Silicon carbide and diamond continue to captivate the imagination of many in the micromachining community. Both materials offer significant advantages, in particular hardness, high stiffness (high Young's modulus), resistance to harsh chemical environments, mechanical stability at high temperature, wide bandgap, and very high thermal conductivity (see Table 2.1). Some micromachining in silicon carbide [8] and diamond has been demonstrated; however, much remains to be studied about both materials and their potential use in MEMS. An important feature of both silicon carbide and diamond is that they exhibit piezoresistive properties. High-temperature pressure sensors in silicon carbide substrates have been developed with stable operation up to about 500°C.

Silicon carbide (SiC) has a number of possible crystal structures, including cubic and hexagonal. Hexagonal crystalline SiC substrates are commercially available, but they are very expensive and are available only in diameters up to 76 mm [9]. Cubic crystalline silicon carbide can be obtained by epitaxial growth directly on silicon (which has the same cubic structure), but the material has a high density of voids and dislocations due to mismatch in lattice spacing. Thin polycrystalline SiC films deposited by chemical vapor deposition can be used as the structural layer for surface micromachining (discussed in Chapter 3), with a sacrificial layer of silicon or silicon dioxide [8]. Because etching SiC is so difficult, alternative methods of forming a pattern, such as selective deposition and using a mold, have been studied. Silicon carbide films have also been used as a coating material for harsh environments.

Diamond is an even lesser-explored material than silicon carbide. Thin synthetic polycrystalline diamond or "diamond-like carbon" films made with thicknesses up to a few microns can be formed using chemical vapor deposition. Diamond has an extremely high ratio of Young's modulus to density, giving vibrating structures made of diamond higher resonant frequencies than similar structures made of other materials. In addition to the properties listed earlier, diamond films are also good field emitters and have received extensive study as a source of electrons for such applications as displays. Etching diamond films is even more difficult than for silicon carbide, so alternative patterning methods such as selective deposition are used [9].

Gallium Arsenide and Other Group III-V Compound Semiconductors

Rather than ponder the utility of gallium arsenide (GaAs) and other group III-V compounds (e.g., InP, AlGaAs, GaN) as alternate substrate materials to silicon, it is perhaps more appropriate to think of micromachining as a set of tools that can provide solutions to issues specific to devices that currently can only be built in these materials, in particular lasers and optical devices. In that regard, micromachining becomes an application-specific toolbox whose main characteristic is to address ways to enable new functions or enhance existing ones.

Micromechanical structures such as springs and bridges have been formed in GaAs by both reactive ion etching [10] and orientation-dependent etching [11] (discussed in Chapter 3). Micromachining has also been used to incorporate structures such as mirrors on the surface of III-V semiconductors to create new devices, including tunable lasers [12]. Moreover, micromachining using GaAs and other group

III-V compound semiconductors is a practical way to integrate RF switches, antennas, and other custom high-frequency components with ultra-high-speed electronic devices for wireless telecommunications.

Polymers

Polymers are long chains of carbon (or sometimes silicon) atoms with various chemical side groups attached to the carbon [13]. If the chains are not crosslinked by covalent bonds, they are able to move relative to each other at elevated temperature under applied stress. Such materials reharden upon cooling and are called thermoplastics. The temperature above which flow readily occurs is the glass *transition temperature,* which varies with the length of the molecules and the type of side groups.

PMMA [poly(methylmethacrylate)], polypropylene, polyvinyl chloride, acrylic, and other thermoplastics are used in sheet form as a substrate for micromachining. Heating above the glass transition temperature enables molding or embossing under pressure from a master for some of these materials (described in Chapter 3). Layers of polycarbonate and acrylic, with channels already formed in their surfaces by hot embossing or conventional machining, have been thermally bonded together for microfluidic systems. In MEMS, thick layers of PMMA have also been spin-coated and used as a photoresist.

Polymer substrates have not been used as much as silicon in micromachining, but have some advantages, perhaps the most important being lower cost. The processing temperatures allowed are much lower than for silicon and many glasses, but suitable fabrication processes have been designed, particularly for biological applications. Polymers are in general less stiff than inorganic materials (see Table 2.1).

Polyimide is a material that is most often used in the form of sheets 7 to 125 μm thick, but can also be spin-coated in films a few micrometers thick. It is sold by DuPont High Performance Films of Circleville, Ohio, under the trade name Kapton®. Polyimide is relatively inert, is a good electrical insulator, and can be exposed to a wide range of temperatures, roughly –250° to +400°C, for at least a short time [14]. In the electronics industry, polyimide has been used as a flexible substrate for printed circuit boards and for hard disk drives. In micromachining, sheets have been laser cut to form microfluidic devices, while spin-on films have been used as resists, sacrificial layers, and a wafer-bonding adhesive.

Other polymers finding application in MEMS include parylenes and silicones. Parylenes are deposited by chemical-vapor deposition to form a conformal coating. There are several forms of parylene due to variations in the chemical structure [15]. Like polyimide, parylenes are fairly inert chemically and form a barrier to the flow of water and other vapors. Silicones are different from most other polymers in that the backbone chain of atoms is silicon rather than carbon. Silicones are very compliant and have been used as the deformable membrane in valves [15], as well as being a common die-attach material in packaging (see Chapter 8).

Shape-Memory Alloys

The shape-memory effect is a unique property of a special class of alloys that return to a predetermined shape when heated above a critical *transition temperature*. The

material "remembers" its original shape after being strained and deformed. The discovery was first made in a gold-cadmium alloy in 1951 but was quickly extended to a broad range of other alloys, including titanium-nickel, copper-aluminum-nickel, iron-nickel and iron-platinum alloys. A basic understanding of the underlying physical principles was established in the 1970s, but extensive research remains ongoing in an effort to develop a thorough theoretical foundation. Nonetheless, the potential applications for shape-memory alloys abound. It has been estimated that upwards of 15,000 patents have been applied for on this topic. Titanium-nickel alloys have been the most widely used of shape-memory alloys because of their relative simple composition and robustness.

An important factor that determines the practical utility of the alloy is its transition temperature. Below this temperature, it has a low yield strength; in other words, it is readily deformed into new permanent shapes. The deformation can be 20 times larger than the elastic deformation. When heated above its transition temperature, the material completely recovers its original (high-temperature) shape through complex changes in its crystal structure. The process generates very large forces, making shape-memory alloys ideal for actuation purposes. By contrast, piezoelectric and electrostatic actuators exert only a fraction of the force available from a shape-memory alloy, but they act much more quickly.

Bulk titanium-nickel alloys in the form of wires and rods are commercially available under the name Nitinol™ [16]. Its transition temperature can be tailored between –100° and 100°C, typically by controlling stoichiometry and impurity concentration. Recently, thin titanium-nickel films with thicknesses up to 50 μm were successfully demonstrated with properties similar to those of Nitinol. Titanium-nickel is a good electrical conductor, with a resistivity of 80 $\mu\Omega\bullet$cm, but a relatively poor thermal conductor, with a conductivity about one tenth that of silicon. Its yield strength is only 100 MPa below its transition temperature but rapidly increases to 560 MPa once heated above it. The Young's modulus shows a similar dependence on temperature; at low temperatures, it is 28 GPa, increasing to 75 GPa above the transition temperature.

Important Material Properties and Physical Effects

The interaction of physical parameters with each other—most notably electricity with mechanical stress, temperature and thermal gradients, magnetic fields, and incident light—yields a multitude of phenomena of great interest to MEMS. We will briefly review in this section three commonly used effects: piezoresistivity, piezoelectricity, and thermoelectricity.

Piezoresistivity

Piezoresistivity is a widely used physical effect and has its name derived from the Greek word *piezein* meaning to apply pressure. Discovered first by Lord Kelvin in 1856, it is the phenomenon by which an electrical resistance changes in response to mechanical stress. The first application of the piezoresistive effect was metal strain gauges to measure strain, from which other parameters such as force, weight, and pressure were inferred (see Figure 2.4). Most the resistance change in metals is due to

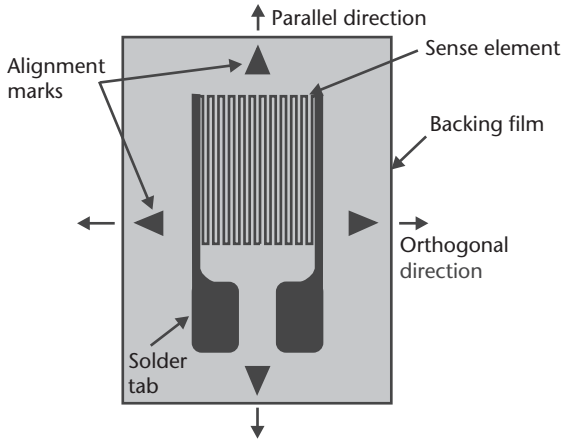


Figure 2.4 A typical thin metal foil strain gauge mounted on a backing film. Stretching of the sense element causes a change in its resistance.

dimensional changes: under stress, the resistor gets longer, narrower, and thinner [17]. C. S. Smith's discovery in 1954 [18] that the piezoresistive effect in silicon and germanium was much greater (by roughly two orders of magnitude) than in metals spurred significant interest. The first pressure sensors based on diffused (impurity-doped) resistors in thin silicon diaphragms were demonstrated in 1969 [19]. The majority of today's commercially available pressure sensors use silicon piezoresistors.

For the physicist at heart, piezoresistivity arises from the deformation of the energy bands as a result of an applied stress. In turn, the deformed bands affect the effective mass and the mobility of electrons and holes, hence modifying resistivity. For the engineer at heart, the fractional change in resistivity, $\Delta\rho/\rho$, is to a first order linearly dependent on $\sigma_{//}$ and $\sigma_{\perp}$, the two stress components parallel and orthogonal to the direction of the resistor, respectively. The direction of the resistor is here defined as that of the current flow. The relationship can be expressed as

$$\Delta\rho/\rho = \pi_{//}\sigma_{//} + \pi_{\perp}\sigma_{\perp}$$

where the proportionality constants, $\pi_{//}$ and $\pi_{\perp}$, are called the parallel and perpendicular piezoresistive coefficients, respectively, and are related to the gauge factor[2] by the Young's modulus of the material. The piezoresistive coefficients depend on crystal orientation and change significantly from one direction to the other (see Table 2.4). They also depend on dopant type (*n*-type versus *p*-type) and concentration. For {100} wafers, the piezoresistive coefficients for *p*-type elements are maximal in the <110> directions and nearly vanish along the <100> directions. In other words, *p*-type piezoresistors must be oriented along the <110> directions to measure stress and thus should be either aligned or perpendicular to the wafer primary flat. Those at 45° with respect to the primary flat (i.e., in the <100> direction), are insensitive to applied tensile stress, which provides an inexpensive

2. The gauge factor, K, is the constant of proportionality relating the fractional change in resistance, $\Delta R/R$, to the applied strain, ε, by the relationship $\Delta R/R = K \cdot \varepsilon$.

way to incorporate stress-independent diffused temperature sensors. The crystal-orientation-dependence of the piezoresistive coefficients takes a more complex function for piezoresistors diffused in {110} wafers, but this dependence fortuitously disappears in {111} wafers. More descriptive details of the underlying physics of piezoresistivity and dependence on crystal orientation can be found in [20, 21].

If we consider *p*-type piezoresistors diffused in {100} wafers and oriented in the <110> direction (parallel or perpendicular to the flat), it is apparent from the positive sign of $\pi_{//}$ in Table 2.4 that the resistance increases with tensile stress applied in the parallel direction, $\sigma_{//}$, as if the piezoresistor itself is being elongated. Furthermore, the negative sign of $\pi_{\perp}$ implies a decrease in resistance with tensile stress orthogonal to the resistor, as if its width is being stretched. In actuality, the stretching or contraction of the resistor are not the cause of the piezoresistive effect, but they make a fortuitous analogy to readily visualize the effect of stress on resistance. This analogy breaks down for *n*-type piezoresistors.

Like many other physical effects, piezoresistivity is a strong function of temperature. For lightly doped silicon (*n*- or *p*-type, 10^{18} cm^{-3}), the temperature coefficient of $\pi_{//}$ and $\pi_{\perp}$ is approximately –0.3% per degree Celsius. It decreases with dopant concentration to about –0.1% per degree Celsius at 8×10^{19} cm^{-3}.

Polysilicon and amorphous silicon also exhibit a strong piezoresistive effect. A wide variety of sensors using polysilicon piezoresistive sense elements have been demonstrated. Clearly, piezoresistive coefficients lose their sensitivity to crystalline direction and become an average over all orientations. Instead, the gauge factor, *K*, relating the fractional change in resistance to strain is often used. Gauge factors in polysilicon and amorphous silicon range typically between –30 and +40, about a third that of single-crystal silicon. The gauge factor decreases quickly as doping concentration exceeds 10^{19} cm^{-3}. However, one advantage of polysilicon over crystalline silicon is its reduced TCR. At doping levels approaching 10^{20} cm^{-3}, the TCR for polycrystalline silicon is approximately 0.04% per degree Celsius compared to 0.14% per degree Celsius for crystalline silicon. The deposition process and the dopant species have been found to even alter the sign of the TCR. For example, emitter-type polysilicon (a special process for depositing heavily doped polysilicon to be used as emitter for bipolar transistors) has a TCR of –0.045% per degree Celsius. Resistors with positive TCR are particularly useful in compensating the negative temperature dependence of piezoresistive sensors.

Piezoelectricity

Certain classes of crystals exhibit the peculiar property of producing an electric field when subjected to an external force. Conversely, they expand or contract in response

Table 2.4 Piezoresistive Coefficients for *n*- and *p*-Type {100} Wafers and Doping Levels Below 10^{18} cm^{-3}

	$\pi_{//}$ (10^{-11} m^2/N)	$\pi_{\perp}$ (10^{-11} m^2/N)	
p-type	7	–1	In <100> direction
	72	–66	In <110> direction
n-type	–102	53	In <100> direction
	–31	–18	In <110> direction

Note: The values decrease precipitously at higher doping concentrations.

to an externally applied voltage. The effect was discovered in quartz by the brothers Pierre and Jacques Curie in 1880 [22]. Its first practical application was in the 1920s when Langevin developed a quartz transmitter and receiver for underwater sound—the first Sonar! Piezoelectric crystals are common in many modern applications (e.g., as clock oscillators in computers and as ringers in cellular telephones). They are attractive for MEMS because they can be used as sensors as well as actuators, and they can be deposited as thin films over standard silicon substrates.

The physical origin of piezoelectricity is explained by charge asymmetry within the primitive unit cell, resulting in the formation of a net electric dipole (see Figure 2.5). Adding up these individual dipoles over the entire crystal gives a net polarization and an effective electric field within the material. Crystal symmetry again plays an important role: Only a crystal that lacks a center of symmetry exhibits piezoelectric properties. A crystal with a center of symmetry, such as a cubic crystal, is not piezoelectric because the net electric dipole within the primitive unit is always vanishing, even in the presence of an externally applied stress (see Figure 2.6). Silicon is not piezoelectric because it is cubic, and, further, the atoms are held together by covalent (not ionic) bonding.

If we consider an ionic or partly ionic crystal lacking a center of symmetry, for example zinc oxide (ZnO), the net electric dipole internal to the primitive unit is zero only in the absence of an externally applied stress. Straining the crystal shifts the relative positions of the positive and negative charges, giving rise to an electric dipole within the primitive unit and a net polarization across the crystal. Conversely, the internal electric dipoles realign themselves in response to an externally applied electric field, causing the atoms to displace and resulting in a measurable crystal deformation. When the temperature exceeds a critical value called the *Curie temperature,* the material loses its piezoelectric characteristics.

The piezoelectric effect is described in terms of piezoelectric charge coefficients, d_{ij}, which relate the static voltage, electric field, or surface charge in the i direction to displacement, applied force, or stress in the j direction. The convention for describing piezoelectrics is that the direction of polarization is the "3" or z direction of the crystal axis, while a direction perpendicular to it is the "1" or x or y direction of the crystal. Hence, piezoelectric charge coefficients are given as d_{33} for both voltage and

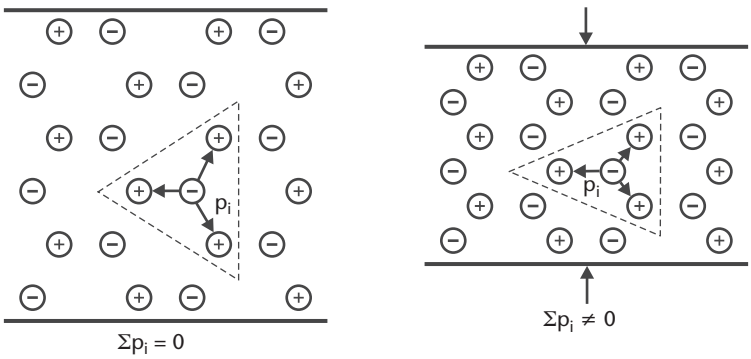


Figure 2.5 Illustration of the piezoelectric effect in a hypothetical two-dimensional crystal. The net electric dipole within the primitive unit of an ionic crystal lacking a center of symmetry does not vanish when external stress is applied. This is the physical origin of piezoelectricity. (*After:* [21].)

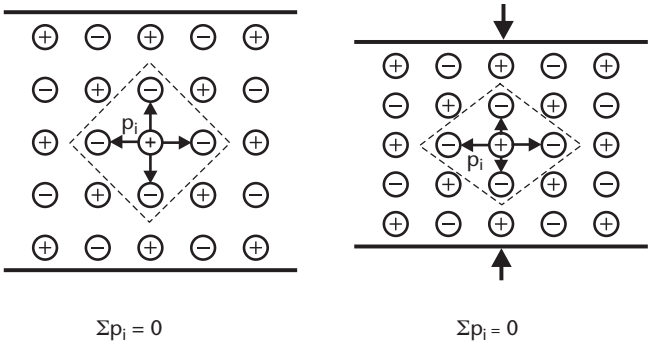


Figure 2.6 Illustration of the vanishing dipole in a two-dimensional lattice. A crystal possessing a center of symmetry is not piezoelectric because the dipoles, p_i, within the primitive unit always cancel each other out. Hence, there is no net polarization within the crystal. An externally applied stress does not alter the center of symmetry. (*After:* [21].)

force along the z axis, and d_{31} for voltage along the z axis but force along the x or y axis. The units of the charge coefficients are C/N, which are the same as m/V. The choice depends on whether the electrical parameter of interest is voltage or charge.

If a voltage, V_a, is applied across the thickness of a piezoelectric crystal (see Figure 2.7), the unconstrained displacements ΔL, ΔW, and Δt along the length, width, and thickness directions, respectively, are given by

$$\Delta L = d_{31} \cdot V_a \cdot L/t \quad \Delta W = d_{31} \cdot V_a \cdot W/t \quad \Delta t = d_{33} \cdot V_a$$

where L and W are the length and width of the plate, respectively, and t is the thickness or separation between the electrodes. In this case, d units of m/V are appropriate. Conversely, if a force, F, is applied along any of the length, width, or thickness directions, a measured voltage, V_m, across the electrodes (in the thickness direction) is given in each of the three cases, respectively, by

$$V_m = d_{31} \cdot F/(\varepsilon \cdot W) \quad V_m = d_{31} \cdot F/(\varepsilon \cdot L) \quad V_m = d_{33} \cdot F \cdot t/(\varepsilon \cdot L \cdot W)$$

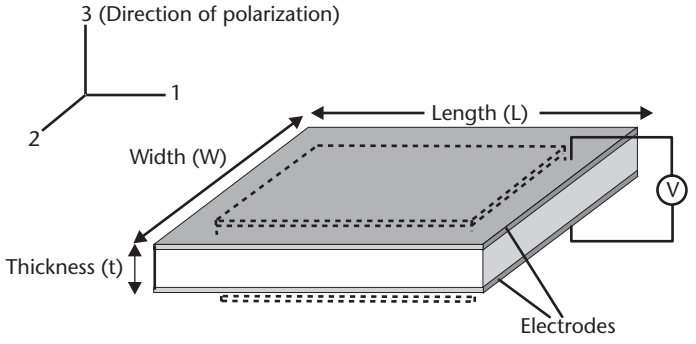


Figure 2.7 An illustration of the piezoelectric effect on a crystalline plate. An applied voltage across the electrodes results in dimensional changes in all three axes (if d_{31} and d_{33} are nonzero). Conversely, an applied force in any of three directions gives rise to a measurable voltage across the electrodes.

where ε is the dielectric permittivity of the material. In this case, d units of C/N are used. The reversibility between strain and voltage makes piezoelectric materials ideal for both sensing and actuation. Further detailed reading on piezoelectricity may be found in [23, 24].

Quartz is a widely used stand-alone piezoelectric material, but there are no available methods to deposit crystalline quartz as a thin film over silicon substrates (see Table 2.5). Piezoelectric ceramics are also common. Lithium niobate ($LiNbO_3$) and barium titanate ($BaTiO_3$) are two well-known examples, but they are also difficult to deposit as thin films. Piezoelectric materials that can be deposited as thin film with relative ease are lead zirconate titanate (PZT)—a ceramic based on solid solutions of lead zirconate ($PbZrO_3$) and lead titanate ($PbTiO_3$)—ZnO, and PVDF. Zinc oxide is typically sputtered and PZT can be either sputtered or deposited in a sol-gel process (Chapter 3 describes the deposition processes in more detail). PVDF is a polymer that can be spun on. All of these deposited films must be *poled* (i.e., polarized by heating above the Curie temperature, then cooling with a large electric field across them) in order to exhibit piezoelectric behavior.

Thermoelectricity

Interactions between electricity and temperature are common and were the subject of extensive studies in the nineteenth century, though the underlying theory was not put in place until early in the twentieth century by Boltzmann. In the absence of a magnetic field, there are three distinct thermoelectric effects: the Seebeck, the Peltier, and the Thomson effects [25]. The Seebeck effect is the most frequently used (e.g., in thermocouples for the measurement of temperature differences). The Peltier effect is used to make thermoelectric coolers (TECs) and refrigerators. The Thomson effect is less known and uncommon in daily applications.

In the Peltier effect, current flow across a junction of two dissimilar materials causes a heat flux, thus cooling one side and heating the other. Mobile wet bars with Peltier refrigerators were touted in 1950s as the newest innovation in home appliances, but their economic viability was quickly jeopardized by the poor energy conversion efficiency. Today, Peltier devices are made of *n*-type and *p*-type bismuth telluride elements and are used to cool high-performance microprocessors, laser diodes, and infrared sensors. Peltier devices have proven to be difficult to implement as micromachined thin-film structures.

Table 2.5 Piezoelectric Coefficients and Other Relevant Properties for a Selected List of Piezoelectric Materials

Material	*Piezoelectric Constant (d_{ij}) (10^{-12} C/N)*	*Relative Permittivity (ε_r)*	*Density (g/cm^3)*	*Young's Modulus (GPa)*	*Acoustic Impedance (10^6 $kg/m^2 \cdot s$)*
Quartz	$d_{33} = 2.31$	4.5	2.65	107	15
Polyvinylidene-fluoride (PVDF)	$d_{31} = 23$ $d_{33} = -33$	12	1.78	3	2.7
$LiNbO_3$	$d_{31} = -4$, $d_{33} = 23$	28	4.6	245	34
BaTiO3	$d_{31} = 78$, $d_{33} = 190$	1,700	5.7		30
PZT	$d_{31} = -171$ $d_{33} = 370$	1,700	7.7	53	30
zinc oxide (ZnO)	$d_{31} = 5.2$, $d_{33} = 246$	1,400	5.7	123	33

In the Seebeck effect, named after the scientist who made the discovery in 1822, a temperature gradient across an element gives rise to a measurable electric field that tends to oppose the charge flow (or electric current) resulting from the temperature imbalance. The measured voltage is, to first order, proportional to the temperature difference with the proportionality constant known as the Seebeck coefficient. While, in theory, a single material is sufficient to measure temperature, in practice, thermocouples employ a junction of two dissimilar materials. The measurable voltage at the leads, ΔV, is the sum of voltages across both legs of the thermocouple. Therefore,

$$\Delta V = \alpha_1 \cdot (T_{cold} - T_{hot}) + \alpha_2 \cdot (T_{hot} - T_{cold}) = (\alpha_2 - \alpha_1) \cdot (T_{hot} - T_{cold})$$

where α_1 and α_2 are the Seebeck coefficients of materials 1 and 2, and, T_{hot} and T_{cold} are the temperatures of the hot and cold sides of the thermocouple, respectively (see Figure 2.8). Alternately, one may use this effect to generate electrical power by maintaining a temperature difference across a junction. Table 2.6 lists Seebeck coefficients for a number of materials.

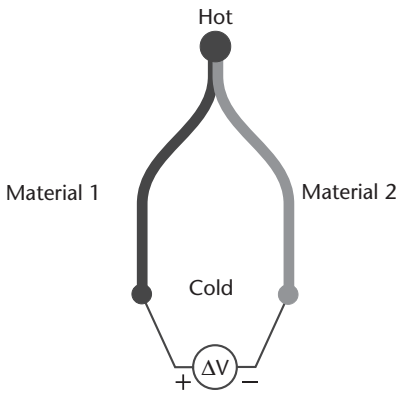


Figure 2.8 The basic structure of a thermocouple using the Seebeck effect. The measured voltage is proportional to the difference in temperature. Thermocouples can be readily implemented on silicon substrates using combinations of thin metal films or polysilicon.

Table 2.6 The Seebeck Coefficients Relative to Platinum for Selected Metals and for *n*- and *p*-Type Polysilicon

	$\mu V/K$		$\mu V/K$
Bi	–73.4	Ag	7.4
Ni	–14.8	Cu	7.6
Pa	–5.7	Zn	7.6
Pt	0	Au	7.8
Ta	3.3	W	11.2
Al	4.2	Mo	14.5
Sn	4.2	*n*-poly (30 Ω/□)	–100
Mg	4.4	*n*-poly (2600 Ω/□)	–450
Ir	6.5	*p*-poly (400 Ω/□)	270

Note: The sheet resistance is given for the 0.38-μm-thick polysilicon films. Polysilicon is an attractive material for the fabrication of thermocouples and thermopiles because of its large Seebeck coefficient.

Summary

The selection of substrate materials for MEMS is very broad, but crystalline silicon is by far the most common choice. Complementing silicon is a host of materials that can be deposited as thin films. These include polysilicon, amorphous silicon, silicon oxides and nitrides, glasses, organic polymers, and a host of metals. Crystallographic planes play an important role in the design and fabrication of silicon-based MEMS and affect some material properties of silicon. Three physical effects commonly used in the operation of micromachined sensors and actuators were introduced: piezoresistivity, piezoelectricity, and thermoelectricity.

References

[1] Ashcroft, N. W., and N. D. Mermin, *Solid State Physics*, Philadelphia, PA: Saunders College, 1976, pp. 91–93.

[2] Kamins, T., *Polycrystalline Silicon for Integrated Circuits*, Boston, MA: Kluwer Academic Publishers, 1988.

[3] Bustillo, J. M., R. T. Howe, and R. S. Muller, "Surface Micromachining for Microelectromechanical Systems," *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1559–1561.

[4] Siethoff, H., "Macroscopic Mechanical Behaviour of Si at High Temperature," in *Properties of Crystalline Silicon*, R. Hull (ed.), London: INSPEC, The Institution of Electrical Engineers, Section 3.6, 1999.

[5] *Properties of Silicon*, EMIS Datareviews Series, No. 4, London: INSPEC, The Institution of Electrical Engineers, 1988.

[6] Lorenz, H., et al., "SU-8: A Low-Cost Negative Resist for MEMS," *Journal of Micromechanics and Microengineering*, Vol. 7, No. 3, September 1997, pp. 121–124.

[7] Gutierrez Monreal, J., and C. M. Mari, "The Use of Polymer Materials As Sensitive Elements in Physical and Chemical Sensors," *Sensors and Actuators*, Vol. 12, 1987, pp. 129–144.

[8] Mehregany, M., et al., "Silicon Carbide MEMS for Harsh Environments," *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1594–1610.

[9] Zorman, C. A., and M. Mehregany, "Materials for Microelectromechanical Systems," in *The MEMS Handbook*, Chapter 15, M. Gad-el-Hak (ed.), Boca Raton, FL: CRC Press, 2002.

[10] Zhang, Z. L., and N. C. MacDonald, "Fabrication of Submicron High-Aspect-Ratio GaAs Actuators," *Journal of Microelectromechanical Systems*, Vol. 2, No. 2, June 1993, pp. 66–73.

[11] Chong, N., T. A. S. Srinivas, and H. Ahmed, "Performance of GaAs Microbridge Thermocouple Infrared Detectors," *Journal of Microelectromechanical Systems*, Vol. 6, No. 2, June 1997, pp. 136–141.

[12] Li, M. Y., et al., "Top-Emitting Micromechanical VCSEL with a 31.6-nm Tuning Range," *IEEE Photonics Technology Letters*, Vol. 10, No. 1, January 1998, pp. 18–20.

[13] Van Vlack, L. H., *Elements of Materials Science and Engineering, 6th Edition*, Reading, MA: Addison-Wesley, 1989, pp. 32–36.

[14] DuPont High Performance Films, "Kapton Data Sheet H-38492-2," Circleville, OH, August 1997.

[15] Yang, X., C. Grosjean, and Y.-C. Tai, "Design, Fabrication, and Testing of Micromachined Silicone Rubber Membrane Valves," *Journal of Microelectromechanical Systems*, Vol. 8, No. 4, December 1999, pp. 393–402.

[16] Rogers, C., "Intelligent Materials," *Scientific American*, Vol. 273, No. 3, September 1995, pp. 154–157.

[17] Kovacs, G. T. A., *Micromachined Transducers Sourcebook*, Boston, MA: WCB/McGraw-Hill, 1998, pp. 211–215.

[18] Smith, C. S., "Piezoresistive Effect in Germanium and Silicon," *Physical Review*, Vol. 94, 1954, pp. 42–49.

[19] Gieles, C. M., "Subminiature Silicon Pressure Sensor Transducer," *Digest IEEE International Solid-State Circuits Conference*, Philadelphia, PA, February 19–21, 1969, pp. 108–109.

[20] Kanda, Y., "A Graphical Representation of the Piezoresistive Coefficients in Silicon," *IEEE Transactions on Electron Devices*, Vol. ED-29, No. 1, 1982, pp. 64–70.

[21] Middelhoek, S., and S. A. Audet, *Silicon Sensors*, San Diego, CA: Academic Press, 1989.

[22] Curie, P., and J. Curie, "Development by Pressure of Polar Electricity in Hemihedral Crystals with Inclined Faces," *Bull. Soc. Min. de France*, Vol. 3, 1880, p. 90.

[23] Cady, W. G., *Piezoelectricity*, New York: Dover, 1964.

[24] Zelenka, J., *Piezoelectric Resonators and their Applications*, Amsterdam, The Netherlands: Elsevier, 1986.

[25] MacDonald, D. K. C., *Thermoelectricity: An Introduction to the Principles*, New York: Wiley, 1962.

Selected Bibliography

Dyos, G. T., and T. Farrell (eds.), *Electrical Resistivity Handbook*, London: Peter Pererginus, 1992.

Hull, R. (ed.), *Properties of Crystalline Silicon*, London: INSPEC, Institution of Electrical Engineers, 1999.

Kittel, C., *Introduction to Solid State Physics*, 6th edition, New York: Wiley, 1986.

Sze, S. M. (ed.), *Semiconductor Sensors*, New York: Wiley, 1994.

Sze, S. M., *Physics of Semiconductor Devices*, 2nd edition, New York: Wiley, 1981.

CHAPTER 3

Processes for Micromachining

> "You will have to brace yourselves for this—not because it is difficult to understand, but because it is absolutely ridiculous: All we do is draw arrows on a piece of paper—that's all!"
>
> —*Richard Feynman, explaining the Theory of Quantum Electrodynamics at the Alix G. Mautner Memorial Lectures, UCLA, 1983.*

This chapter presents methods used in the fabrication of MEMS. Many are borrowed from the integrated-circuit industry, in addition to others developed specifically for silicon micromachining. There is no doubt that the use of process equipment and the corresponding portfolio of fabrication processes initially developed for the semiconductor industry has given the burgeoning MEMS industry the impetus it needs to overcome the massive infrastructure requirements. For example, lithographic tools used in micromachining are oftentimes from previous generations of equipment designed for the fabrication of electronic integrated circuits. The equipment's performance is sufficient to meet the requirements of micromachining, but its price is substantially discounted. A few specialized processes, such as anisotropic chemical wet etching, wafer bonding, deep reactive ion etching, sacrificial etching, and critical-point drying, emerged over the years within the MEMS community and remain limited to micromachining in their application.

From a simplistic perspective, micromachining bears a similarity to conventional machining in the sense that the objective is to precisely define arbitrary features in or on a block of material. There are, however, distinct differences. Micromachining is a parallel (batch) process in which dozens to tens of thousands of identical elements are fabricated simultaneously on the same wafer. Furthermore, in some processes, dozens of wafers are processed at the same time. Another key difference is the minimum feature dimension—on the order of one micrometer—which is an order of magnitude smaller than what can be achieved using conventional machining.

Silicon micromachining combines adding layers of material over a silicon wafer with etching (selectively removing material) precise patterns in these layers or in the underlying substrate. The implementation is based on a broad portfolio of fabrication processes, including material deposition, patterning, and etching techniques. Lithography plays a significant role in the delineation of accurate and precise patterns. These are the tools of MEMS (see Figure 3.1).

We divide the toolbox into three major categories: basic, advanced, and nonlithographic processes. The basic process tools are well-established methods and are usually available at major foundry facilities. The advanced process tools are unique in their nature and are normally limited to a few specialized facilities. For example,

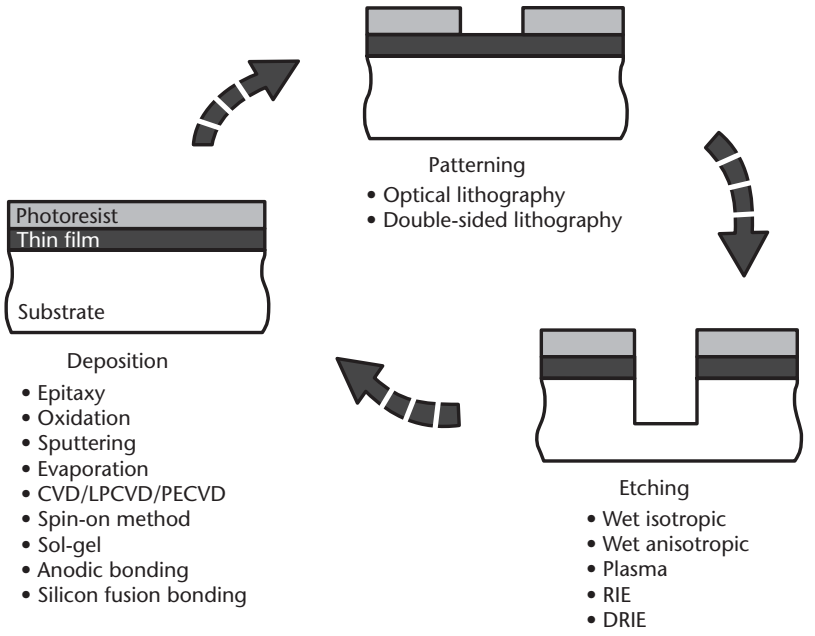


Figure 3.1 Illustration of the basic process flow in micromachining: Layers are deposited; photoresist is lithographically patterned and then used as a mask to etch the underlying materials. The process is repeated until completion of the microstructure.

very few sites offer LIGA[1], a micromachining process using electroplating and molding. The nonlithographic processes are more conventional means of producing microstructures, which may be combined with other processes to produce a final MEMS product.

Basic Process Tools

Epitaxy, sputtering, evaporation, chemical-vapor deposition, and spin-on methods are common techniques used to deposit uniform layers of semiconductors, metals, insulators, and polymers. Lithography is a photographic process for printing images onto a layer of photosensitive polymer (photoresist) that is subsequently used as a protective mask against etching. Wet and dry etching, including deep reactive ion etching, form the essential process base to selectively remove material. The following sections describe the fundamentals of the basic process tools.

Epitaxy

Epitaxy is a deposition method to grow a crystalline silicon layer over a silicon wafer, but with a differing dopant type and concentration. The epitaxial layer is typically 1 to 20 μm thick. It exhibits the same crystal orientation as the underlying crystalline substrate, except when grown over an amorphous material (e.g., a layer

1. LIGA is a German acronym for lithographie, galvanoformung, und abformung, meaning lithography, electroplating, and molding.

of silicon dioxide), it is polycrystalline. Epitaxy is a widely used step in the fabrication of CMOS circuits and has proven efficient in forming wafer-scale *p-n* junctions for controlled electrochemical etching (described later).

The growth occurs in a vapor-phase chemical-deposition reactor from the dissociation or hydrogen reduction at high temperature (>800°C) of a silicon-containing source gas. Common source gases are silane (SiH_4), dichlorosilane (SiH_2Cl_2), or silicon tetrachloride ($SiCl_4$). Nominal growth rates are between 0.2 and 4 μm/min, depending on the source gas and the growth temperature. Impurity dopants are simultaneously incorporated during growth by the dissociation of a dopant source gas in the same reactor. Arsine (AsH_3) and phosphine (PH_3), two extremely toxic gases, are used for arsenic and phosphorous (*n*-type) doping, respectively; diborane (B_2H_6) is used for boron (*p*-type) doping.

Epitaxy can be used to grow crystalline silicon on other types of crystalline substrates such as sapphire (Al_2O_3). The process is called heteroepitaxy to indicate the difference in materials. Silicon-on-sapphire (SOS) wafers are available from a number of vendors and are effective in applications where an insulating or a transparent substrate is required. The lattice mismatch between the sapphire and silicon crystals limits the thickness of the silicon to about one micrometer. Thicker silicon films suffer from high defect densities and degraded electronic performance.

Oxidation

High-quality amorphous silicon dioxide is obtained by oxidizing silicon in either dry oxygen or in steam at elevated temperatures (850°–1,150°C). Oxidation mechanisms have been extensively studied and are well understood. Charts showing final oxide thickness as function of temperature, oxidizing environment, and time are widely available [1].

Thermal oxidation of silicon generates compressive stress in the silicon dioxide film. There are two reasons for the stress: Silicon dioxide molecules take more volume than silicon atoms, and there is a mismatch between the coefficients of thermal expansion of silicon and silicon dioxide. The compressive stress depends on the total thickness of the silicon dioxide layer and can reach hundreds of MPa. As a result, thermally grown oxide films cause bowing of the underlying substrate. Moreover, freestanding membranes and suspended cantilevers made of thermally grown silicon oxide tend to warp or curl due to stress variation through the thickness of the film.

Sputter Deposition

In sputter deposition, a *target* made of a material to be deposited is physically bombarded by a flux of inert-gas ions (usually argon) in a vacuum chamber at a pressure of 0.1–10 Pa. Atoms or molecules from the target are ejected and deposited onto the wafer. There are several general classes of sputter tools differing by the ion excitation mechanism. In *direct-current (dc) glow discharge*, suitable only for electrically conducting materials, the inert-gas ions are accelerated in a dc electric field between the target and the wafer. In *planar RF*, the target and the wafer form two parallel plates with RF excitation applied to the target. In *ion-beam deposition* (also known as *ion milling*), ions are generated in a remote plasma, then accelerated at the target.

RF planar sputtering and ion-beam methods work for the deposition of both conducting and insulating materials, such as silicon dioxide. In *planar* and *cylindrical magnetron* sputtering, an externally applied magnetic field increases the ion density near the target, thus raising the deposition rates. Typical deposition rates are 0.1–0.3 μm/min, and can be as high as 1 μm/min for aluminum in certain sputtering tools.

Nearly any inorganic material can be sputtered. Sputtering is a favored method in the MEMS community for the deposition at low temperatures (<150°C) of thin metal films such as aluminum, titanium, chromium, platinum, palladium, tungsten, Al/Si and Ti/W alloys, amorphous silicon, insulators including glass, and piezoelectric ceramics (e.g., PZT and ZnO). In a variation known as reactive sputtering, a reactive gas such as nitrogen or oxygen is added during the sputtering of a metal to form compounds such as titanium nitride or titanium dioxide.

The directional randomness of the sputtering process, provided that the target size is larger than the wafer, results in good *step coverage*—the uniformity of the thin film over a geometrical step—though some thinning occurs near corners. The deposited film has a very fine granular structure and is usually under stress [2]. The stress levels vary with the sputter power and chamber pressure during deposition, with tensile stress occurring at lower power and higher pressure, and compressive stress occurring at higher power and lower pressure. The transition between the compressive and tensile regimes is often sharp (over a few tenths of a Pa), making the crossover—an ideal point for zero-stress deposition—difficult to control. Heating the substrate during deposition is sometimes used to reduce film stress.

Many metals, particularly inert ones such as gold, silver, and platinum, do not adhere well to silicon, silicon dioxide, or silicon nitride, peeling off immediately after deposition or during later handling. A thin (5- to 20-nm) adhesion layer, which bonds to both the underlying material and the metal over it, enables the inert metal to stick. The most common adhesion layers are Cr, Ti, and Ti/W alloy. The inert metal must be deposited on the adhesion layer without breaking the vacuum, as oxygen in the air would immediately oxidize the adhesion layer, rendering it useless.

Evaporation

Evaporation involves the heating of a source material to a high temperature, generating a vapor that condenses on a substrate to form a film. Nearly any element (e.g., Al, Si, Ti, Au), including many high-melting-point (refractory) metals and compounds (e.g., Cr, Mo, Ta, Pd, Pt, Ni/Cr, Al_2O_3), can be evaporated. Deposited films comprised of more than one element may not have the same composition as the source material because the evaporation rates may not correspond to the stoichiometry of the source.

Evaporation is performed in a vacuum chamber with the background pressure typically below 10^{-4} Pa to avoid contaminating the film. Target heating can be done resistively by passing an electrical current through a tungsten filament, strip, or boat holding the desired material. Heating can alternatively be done by scanning a high-voltage (e.g., 10-kV) electron beam (e-beam) over the source material. In this case, the carrier is usually made of tungsten, graphite, alumina, or copper (copper is an excellent thermal conductor, but it can only be used if it is not wetted by the molten source). Resistive evaporation is simple but can result in spreading impurities or other contaminants present in the filament. E-beam evaporation, by contrast, can

provide better-quality films and slightly higher deposition rates (5–100 nm/min), but the deposition system is more complex, requiring water cooling of the target and shielding from x-rays generated when the energetic electrons strike the target. Furthermore, radiation that penetrates the surface of the silicon substrate during the deposition process can damage the crystal and degrade the characteristics of electronic circuits.

Evaporation is a directional deposition process from a relatively small source. This results in the majority of material particles being deposited at a specific angle to the substrate, causing poor step coverage and leaving corners and sidewalls exposed. This is generally an undesirable effect if thin film continuity is desired (e.g., when the metal is an electrical interconnect). Rotating the substrate to face the source at different angles during deposition reduces the effect. In some cases, however, shadowing can be used deliberately to selectively deposit material on one side of a step or a trench but not the other.

Thin films deposited by evaporation tend to exhibit tensile stress, increasing with higher material melting point. Evaporated niobium and platinum films, for example, can have tensile stress in excess of 1 GPa, sufficient to cause curling of the wafer or even peeling. As with sputtering, an adhesion layer must be used with many metals.

Chemical-Vapor Deposition

Chemical-vapor deposition (CVD) works on the principle of initiating a surface chemical reaction in a controlled atmosphere, resulting in the deposition of a reacted species on a heated substrate. In contrast to sputtering, CVD is a high-temperature process, usually performed above 300°C. The field of CVD has grown substantially, driven by the demand within the semiconductor industry for high-quality, thin dielectric and metal films for multilayer electrical interconnects. Common thin films deposited by CVD include polysilicon, silicon oxides and nitrides, tungsten, titanium and tantalum as well as their nitrides, and, most recently, copper and low-permittivity dielectric insulators ($\varepsilon_r < 3$). The latter two are becoming workhorse materials for very-high-speed electrical interconnects in integrated circuits. The deposition of polysilicon, silicon oxides, and nitrides is routine within the MEMS industry.

Chemical vapor deposition processes are categorized as *atmospheric-pressure* (referred to as APCVD), or *low-pressure* (LPCVD), or *plasma-enhanced* (PECVD),[2] which also encompasses high-density plasma (HDP-CVD). APCVD and LPCVD methods operate at rather elevated temperatures (400°–800°C). In PECVD and HDP-CVD, the substrate temperature is typically near 300°C, though the plasma deposition of silicon nitrides at room temperature is feasible. The effect of deposition parameters on the characteristics of the thin film is significant, especially for silicon oxides and nitrides. Substrate temperature, gas flows, presence of dopants,

2. Energetic electrons excited in a high-frequency electromagnetic field collide with gas molecules to form ions and reactive neutral species. The mixture of electrons, ions, and neutrals is called plasma and constitutes a phase of matter distinct from solids, liquids, or gases. Plasma-phase operation increases the density of ions and neutral species that can participate in a chemical reaction, be it deposition or etching, and thus can accelerate the reaction rate.

and pressure are important process variables for all types of CVD. Power and plasma excitation RF frequency are also important for PECVD.

Deposition of Polysilicon

Chemical-vapor deposition processes allow the deposition of polysilicon as a thin film on a silicon substrate. The film thickness can range between a few tens of nanometers to several micrometers. Structures with several layers of polysilicon are feasible. The ease of depositing polysilicon, a material sharing many of the properties of bulk silicon, makes it an extremely attractive material in surface micromachining (described later).

Polysilicon is deposited by the pyrolysis of silane (SiH_4) to silicon and hydrogen in a LPCVD reactor. Deposition from silane in a low-temperature PECVD reactor is also possible but results in amorphous silicon. The deposition temperature in LPCVD, typically between 550° and 700°C, affects the granular structure of the film. Below about 600°C, the thin film is completely amorphous; above about 630°C, it exhibits a crystalline grain structure. The deposition rate varies from approximately 6 nm/min at 620°C up to 70 nm/min at 700°C. Partial pressure and flow rate of the silane gas also affect the deposition rate.

Generally speaking, LPCVD polysilicon films conform well to the underlying topography on the wafer, showing good step coverage. In deep trenches with aspect ratios (ratio of depth to width) in excess of 10, some thinning of the film occurs on the sidewalls, but that has not limited using polysilicon to fill trenches as deep as 500 μm.

Polysilicon can be doped during deposition—known as *in situ* doping—by introducing dopant source gases, in particular arsine or phosphine for *n*-type doping and diborane for *p*-type doping. Arsine and phosphine greatly decrease the deposition rate (to about one third that of undoped polysilicon), whereas diborane increases it. The dopant concentration in *in-situ* doped films is normally very high ($\sim 10^{20}$ cm^{-3}), but the film resistivity remains in the range of 1 to 10 m$\Omega \bullet$cm because of the low mobility of electrons or holes.

Intrinsic stresses in as-deposited doped polysilicon films can be large ($>$500 MPa) and either tensile or compressive, depending on the deposition temperature. Furthermore, there is normally a stress gradient through the thickness of the film, which results in curling of released micromechanical structures. Annealing at 900°C or above causes stress relaxation through structural changes in grain boundaries and a reduction in stress to levels ($<$50 MPa) and stress gradient generally deemed acceptable for micromachined structures.

Deposition of Silicon Dioxide

Silicon dioxide is deposited below 500°C by reacting silane and oxygen in an APCVD, LPCVD, or PECVD reactor. Due to the low temperature compared to thermally grown oxide, this is known as low-temperature oxide (LTO). The optional addition of phosphine or diborane dopes the silicon oxide with phosphorus or boron, respectively. Films doped with phosphorus are often referred to as phosphosilicate glass (PSG); those doped with phosphorus and boron are known as borophosphosilicate glass (BPSG). When annealed at temperatures near 1,000°C,

both PSG and BPSG soften and flow to conform with the underlying surface topography and to improve step coverage. LTO films are used for passivation coatings over aluminum, but the deposition temperature must remain below about 400°C to prevent degradation of the metal.

Silicon dioxide can also be deposited at temperatures between 650° and 750°C in a LPCVD reactor by the pyrolysis of tetraethoxysilane [$Si(OC_2H_4)_4$], also known as TEOS. Silicon dioxide layers deposited from a TEOS source exhibit excellent uniformity and step coverage, but the high temperature process precludes their use over aluminum.

A third, but less common, method to deposit silicon dioxide involves reacting dichlorosilane ($SiCl_2H_2$) with nitrous oxide (N_2O) in a LPCVD reactor at temperatures near 900°C. Film properties and uniformity are excellent, but its use is limited to depositing insulating layers over polysilicon.

As is the case for the LPCVD of polysilicon, deposition rates for silicon dioxide increase with temperature. A typical LTO deposition rate at low pressure is 25 nm/min at 400°C, rising to 150 nm/min at atmospheric pressure and 450°C; the deposition rate using TEOS varies from 5 nm/min at 650°C up to 50 nm/min at 750°C.

Deposited silicon dioxide films are amorphous with a structure similar to fused silica. Heat treatment (annealing) at elevated temperatures (600°–1,000°C) results in the outgassing of hydrogen incorporated in the film and a slight increase in density, but no change in the amorphous structure. This process is called densification.

Silicon dioxide deposited using CVD methods is very useful as a dielectric insulator between layers of metal or as a sacrificial layer (etched using hydrofluoric acid) in surface micromachining. However, its electric properties are inferior to those of thermally grown silicon dioxide. For example, dielectric strength of CVD silicon oxides can be half that of thermally grown silicon dioxide. It is no coincidence that gate insulators for CMOS transistors are made of the latter type. In general, CVD silicon oxides are under compressive stress (100–300 MPa). The stress cannot be controlled except when PECVD is used.

Deposition of Silicon Nitrides

Silicon nitride is common in the semiconductor industry for the passivation of electronic devices because it forms an excellent protective barrier against the diffusion of water and sodium ions. In micromachining, LPCVD silicon nitride films are effective as masks for the selective etching of silicon in alkaline solutions, such as potassium hydroxide. Silicon nitride has also been used as a structural material.

Stoichiometric silicon nitride (Si_3N_4) is deposited at atmospheric pressure by reacting silane (SiH_4) and ammonia (NH_3), or at low pressure by reacting dichlorosilane ($SiCl_2H_2$) and ammonia. The deposition temperature for either method is between 700° and 900°C. Both reactions generate hydrogen as a byproduct, some of which is incorporated in the deposited film. CVD and LPCVD silicon nitride films generally exhibit large tensile stresses approaching 1,000 MPa. However, if LPCVD silicon nitride is deposited at 800°–850°C and is silicon-rich (an excess of silicon in the film) due to a greatly increased dichlorosilane flow rate, the stress can be below 100 MPa—a level acceptable for most micromachining applications.

For deposition below 400°C, nonstoichiometric silicon nitride (Si_xN_y) is obtained by reacting silane with ammonia or nitrogen in a PECVD chamber. Hydrogen is also a byproduct of this reaction and is incorporated in elevated concentrations (20%–25%) in the film. The refractive index is an indirect measure of the stoichiometry of the silicon nitride film. The refractive index for stoichiometric LPCVD silicon nitride is 2.01 and ranges between 1.8 and 2.5 for PECVD films. A high value in the range is indicative of excess silicon, and a low value generally represents an excess of nitrogen.

One of the key advantages of PECVD nitride is the ability to control stress during deposition. Silicon nitride deposited at a plasma excitation frequency of 13.56 MHz exhibits tensile stress of about 400 MPa, whereas a film deposited at a frequency of 50 kHz has a compressive stress of 200 MPa. By alternating frequencies during deposition, one may obtain lower-stress films.

Spin-On Methods

Spin-on is a process to put down layers of dielectric insulators and organic materials. Unlike the methods described earlier, the equipment is simple, requiring a variable-speed spinning table with appropriate safety screens. A nozzle dispenses the material as a liquid solution in the center of the wafer. Spinning the substrate at speeds of 500 to 5,000 rpm for 30 to 60 seconds spreads the material to a uniform thickness.

Photoresists and polyimides are common organic materials that can be spun on a wafer with thicknesses typically between 0.5 and 20 μm, though some special-purpose resists such as epoxy-based SU-8 can exceed 200 μm. The organic polymer is normally in suspension in a solvent solution; subsequent baking causes the solvent to evaporate, forming a firm film.

Thick (5–100 μm) spin-on glass (SOG) has the ability to uniformly coat surfaces and smooth out underlying topographical variations, effectively *planarizing* surface features. Thin (0.1–0.5 μm) SOG was heavily investigated in the integrated circuit industry as an interlayer dielectric between metals for high-speed electrical interconnects; however, its electrical properties are considered poor compared to thermal or CVD silicon oxides. Spin-on glass is commercially available in different forms, commonly siloxane- or silicate-based. The latter type allows water absorption into the film, resulting in a higher relative dielectric constant and a tendency to crack. After deposition, the layer is typically densified at a temperature between 300° and 500°C. Measured film stress is approximately 200 MPa in tension but decreases substantially with increasing anneal temperatures.

Lithography

Lithography involves three sequential steps:

- Application of photoresist (or simply "resist"), which is a photosensitive emulsion layer;
- Optical exposure to print an image of the mask onto the resist;
- Immersion in an aqueous developer solution to dissolve the exposed resist and render visible the latent image.

The mask itself consists of a patterned opaque chromium (the most common), emulsion, or iron oxide layer on a transparent fused-quartz or soda-lime glass substrate. The pattern layout is generated using a computer-aided design (CAD) tool and transferred into the opaque layer at a specialized mask-making facility, often by electron-beam or laser-beam writing. A complete microfabrication process normally involves several lithographic operations with different masks.

Positive photoresist is an organic resin material containing a sensitizer. It is spin-coated on the wafer with a typical thickness between 0.5 μm and 10 μm. As mentioned earlier, special types of resists can be spun to thicknesses of over 200 μm, but the large thickness poses significant challenges to exposing and defining features below 25 μm in size. The sensitizer prevents the dissolution of unexposed resist during immersion in the developer solution. Exposure to light in the 200- to 450-nm range (ultraviolet to blue) breaks down the sensitizer, causing exposed regions to immediately dissolve in developer solution. The exact opposite process happens in negative resists—exposed areas remain and unexposed areas dissolve in the developer.

Optical exposure can be accomplished in one of three different modes: contact, proximity, or projection. In contact lithography, the mask touches the wafer. This normally shortens the life of the mask and leaves undesired photoresist residue on the wafer and the mask. In proximity mode, the mask is brought to within 25 to 50 μm of the resist surface. By contrast, projection lithography projects an image of the mask onto the wafer through complex optics (see Figure 3.2).

Resolution, defined as the minimum feature the optical system can resolve, is seldom a limitation for micromachining applications. For proximity systems, it is limited by Fresnel diffraction to a minimum of about 5 μm, and in contact systems, it is approximately 1 to 2 μm. For projection systems, it is given by $0.5\times \lambda/NA$ where λ is the wavelength (~ 400 nm) and NA is the numerical aperture of the optics (~ 0.25 for steppers used in MEMS). Resolution in projection lithography is

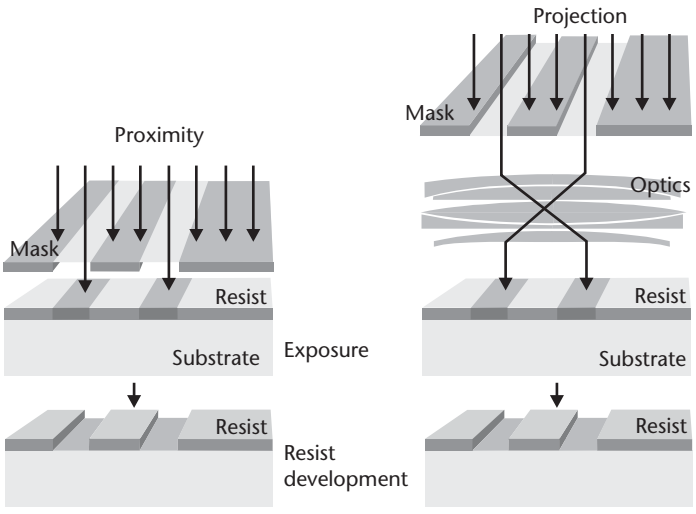


Figure 3.2 An illustration of proximity and projection lithography. In proximity mode, the mask is within 25 to 50 μm of the resist. Fresnel diffraction limits the resolution and minimum feature size to ~ 5 μm. In projection mode, complex optics image the mask onto the resist. The resolution is routinely better than one micrometer. Subsequent development delineates the features in the resist.

routinely better than one micrometer. Depth of focus, however, is a more severe constraint on lithography, especially in light of the need to expose thick resist or accommodate geometrical height variations across the wafer. Depth of focus for contact and proximity systems is poor, also limited by Fresnel diffraction. In projection systems, the image plane can be moved by adjusting the focus settings, but once it is fixed, the depth of focus about that plane is limited to $\pm 0.5 \times \lambda / NA^2$. Depth of focus is typically limited to few microns.

Projection lithography is clearly a superior approach, but an optical projection system can cost significantly more than a proximity or contact system. Long-term cost of ownership plays a critical role in the decision to acquire a particular lithographic tool.

While resolution of most lithographic systems is not a limitation for MEMS, lithography can be challenging depending on the nature of the application; examples include exposure of thick resist, topographical height variations, front to back side pattern alignment, and large fields of view.

Thick Resist

Patterned thick resist is normally used as a protective masking layer for the etching of deep structures and can also be used as a template for the electroplating of metal microstructures. Coating substrates with thick resist is achieved either by multiple spin-coating applications (up to a total of 20 μm) or by spinning special viscous resist solutions at slower speeds (up to 100 μm). Maintaining thickness control and uniformity across the wafer becomes difficult with increasing resist thickness.

Exposing resist thicker than 5 μm often degrades the minimum resolvable feature size due to the limited depth of focus of the exposure tool—different depths within the resist will be imaged differently. The net result is a sloping of the resist profile in the exposed region. As a general guideline, the maximum aspect ratio (ratio of resist thickness to minimum feature dimension) is approximately three—in other words, the minimum achievable feature size (e.g., line width or spacing between lines) is larger than one third of the resist thickness. This limitation may be overcome using special exposure methods, but their value in a manufacturing environment remains questionable.

Topographical Height Variations

Changes in topography on the surface of the wafer, such as deep cavities and trenches, are common in MEMS and pose challenges to both resist spinning and imaging. For cavities deeper than about 10 μm, thinning of the resist at convex corners and accumulation inside the cavity create problems with exposure and with leaving insufficient resist thickness during etches (see Figure 3.3). Two recent developments targeting resist coating of severe topography are spray-on resist and electroplated resist.

Exposing a pattern on a surface with height variations in excess of 10 μm is also a difficult task because of the limited depth of focus. Contact and proximity tools are not suitable for this task unless a significant loss of resolution is tolerable. Under certain circumstances where the number of height levels is limited (say, less than three), one may use a projection lithography tool to perform an exposure with a

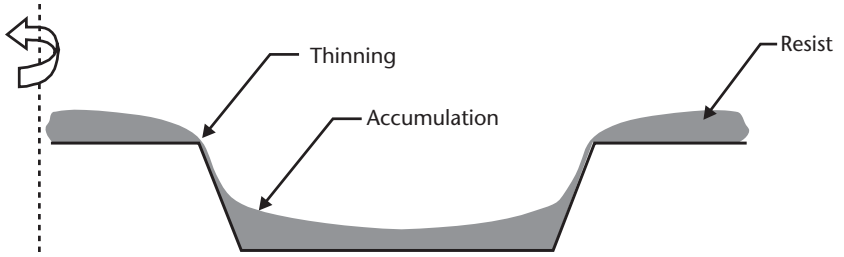


Figure 3.3 Undesirable effects of spin-coating resist on a surface with severe topographical height variations. The resist is thin on corners and accumulates in the cavity.

corresponding focus adjustment at each of these height levels. Naturally, this is costly because the number of masks and exposures increases linearly with the number of height levels.

Double-Sided Lithography

Often, lithographic patterns on both sides of a wafer need to be aligned with respect to each other with high accuracy. For example, the fabrication of a commercial pressure sensor entails forming on the front side of the wafer piezoresistive sense elements that are aligned to the edges of a cavity on the back side of the wafer. Different methods of front-to-back side alignment, also known as double-sided alignment, have been incorporated in commercially available tools. Wafers polished on both sides should be used to minimize light scattering during lithography.

Several companies, including SÜSS MicroTec (formerly Karl Süss) of Munich, Germany, EV Group (formerly Electronic Visions) of Schärding, Austria, OAI (formerly Optical Associates) of San Jose, California, and Ultratech, Inc., of San Jose, California, provide equipment capable of double-sided alignment and exposure. The operation of the SÜSS MA-6 system uses a patented scheme to align crosshair marks on the mask to crosshair marks on the back side of the wafer (see Figure 3.4). First, the alignment marks on the mechanically clamped mask are viewed from below by a set of dual objectives, and an image is electronically stored. The wafer is then loaded with the back side alignment marks facing the microscope objectives and positioned such that these marks are aligned to the electronically stored image. After alignment, exposure of the mask onto the front side of the wafer is completed in proximity or contact mode. A typical registration error (or misalignment) is less than 2 μm.

Large Field of View

The field of view is the extent of the area that is exposed at any one time on the wafer. In proximity and contact lithography, it covers the entire wafer. In projection systems, the field of view is often less than 1×1 cm^2. The entire wafer is exposed by stepping the small field of view across in a two-dimensional array, hence the *stepper* appellation. In some applications, the device structure may span dimensions exceeding the field of view. A remedy to this is called *field stitching*, in which two or more different fields are exposed sequentially, with the edges of the fields overlapping.

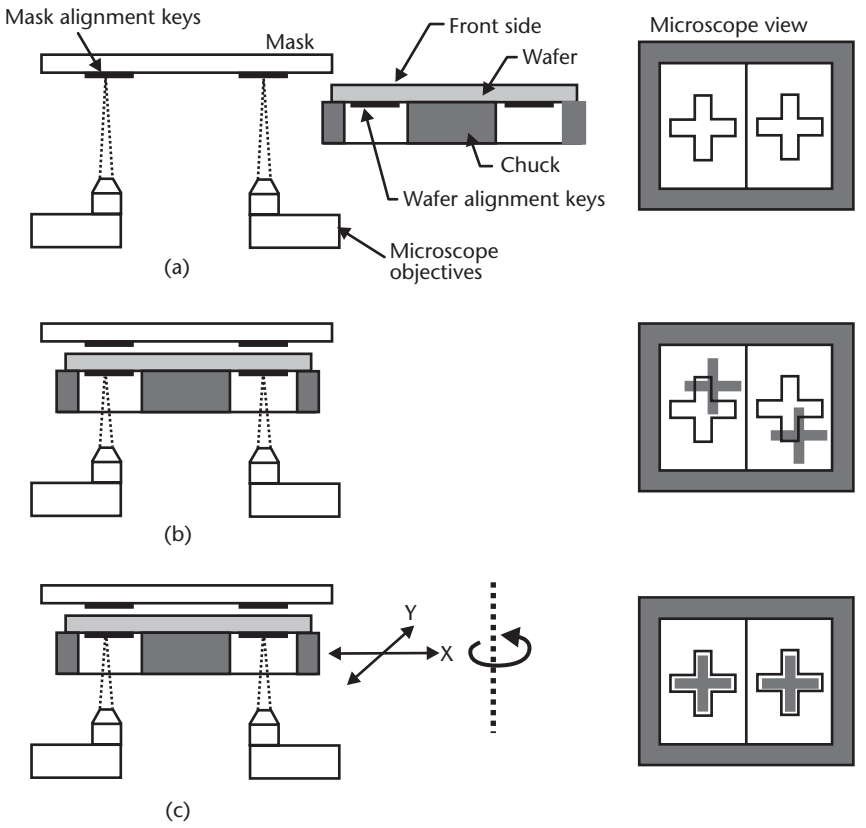


Figure 3.4 Double-sided alignment scheme for the SÜSS MA-6 alignment system: (a) the image of mask alignment marks is electronically stored; (b) the alignment marks on the back side of the wafer are brought in focus; and (c) the position of the wafer is adjusted by translation and rotation to align the marks to the stored image. The right-hand side illustrates the view on the computer screen as the targets are brought into alignment. (*After:* product technical sheet of SÜSS MicroTec of Munich, Germany.)

Etching

In etching, the objective is to selectively remove material using imaged photoresist as a masking template. The pattern can be etched directly into the silicon substrate or into a thin film, which may in turn be used as a mask for subsequent etches. For a successful etch, there must be sufficient selectivity (etch-rate ratio) between the material being etched and the masking material. Etch processes for MEMS fabrication deviate from traditional etch processes for the integrated circuit industry and remain to a large extent an art.

Etching thin films is relatively easier than etching bulk silicon. Table 3.1 provides a list of wet and dry (usually plasma) etchants commonly used for metal and dielectric films.

Deep etching of silicon lies at the core of what is often termed *bulk micromachining*. No ideal silicon etch method exists, leaving process engineers with techniques suitable for some applications but not others. Distinctions are made on the basis of isotropy, etch medium, and selectivity of the etch to other materials.

Table 3.1 Wet and Dry Etchants of Thin Metal Films and Dielectric Insulators

	Wet Etchants (Aqueous Solutions)	*Etch Rate (nm/min)*	*Dry Etching Gases (Plasma or Vapor Phase)*	*Etch Rate (nm/min)*
Thermal silicon dioxide	HF	2,300	CHF_3+ O_2	50–150
	5 NH_4F:1 HF (buffered HF)	100	CHF_3+ CF_4 + He	250–600
			HF vapor (no plasma)	66
LPCVD silicon nitride	Hot H_3PO_4	5	SF_6	150–250
			CHF_3 + CF_4 + He	200–600
Aluminum	Warm H_3PO_4:HNO_3:CH_3COOH	530	Cl_2 + $SiCl_4$	100–150
	HF	4	Cl_2+ BCl_3 +$CHCl_3$	200–600
Gold	KI:I_2	660		
Titanium	HF:H_2O_2	110–880	SF_6	100–150
Tungsten	Warm H_2O_2	150	SF_6	300–400
	$K_3Fe(CN)_6$:KOH:KH_2PO_4	34		
Chromium	$Ce(NH_4)_2(NO_3)_6$:CH_3COOH	93	Cl_2	5
Photoresist	Hot H_2SO_4:H_2O_2	>100,000	O_2	350
	CH_3COOH_3 (acetone)	>100,000		

(*After:* [3, 4].)

Isotropic etchants etch uniformly in all directions, resulting in rounded cross-sectional features. By contrast, anisotropic etchants etch in some directions preferentially over others, resulting in trenches or cavities delineated by flat and well-defined surfaces, which need not be perpendicular to the surface of the wafer (see Figure 3.5). The etch medium (wet versus dry) plays a role in selecting a suitable etch method. Wet etchants in aqueous solution offer the advantage of low-cost batch fabrication—25 to 50 100-mm-diameter wafers can be etched simultaneously—and can be either of the isotropic or anisotropic type. Dry etching involves the use of reactant gases, usually in a low-pressure plasma, but nonplasma gas-phase etching is also used to a small degree. It can be isotropic or vertical. The equipment for dry etching is specialized and requires the plumbing of ultra-clean pipes to bring high-purity reactant gases into the vacuum chamber.

Isotropic Wet Etching

The most common group of silicon isotropic wet etchants is *HNA*, also known as *iso etch* and *poly etch* because of its use in the early days of the integrated circuit industry as an etchant for polysilicon. It is a mixture of hydrofluoric (HF), nitric (HNO_3), and acetic (CH_3COOH) acids, although water may replace the acetic acid. In the chemical reaction, the nitric acid oxidizes silicon, which is then etched by the hydrofluoric acid. The etch rate of silicon can vary from 0.1 to over 100 μm/min depending on the proportion of the acids in the mixture. Etch uniformity is normally difficult to control but is improved by stirring.

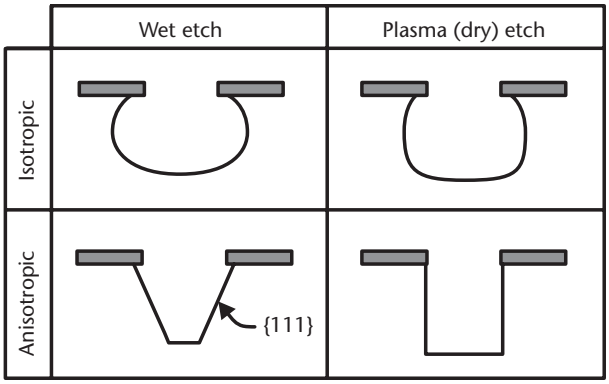


Figure 3.5 Schematic illustration of cross-sectional trench profiles resulting from four different types of etch methods.

Anisotropic Wet Etching

Anisotropic wet etchants are also known as orientation-dependent etchants (ODEs) because their etch rates depend on the crystallographic direction. The list of anisotropic wet etchants includes the hydroxides of alkali metals (e.g., NaOH, KOH, CsOH), simple and quaternary ammonium hydroxides (e.g., NH_4OH, $N(CH_3)_4OH$), and ethylenediamine mixed with pyrochatechol (EDP) in water [5]. The solutions are typically heated to 70°–100°C. A comparison of various silicon etchants is given in Table 3.2.

KOH is by far the most common ODE. Etch rates are typically given in the [100] direction, corresponding to the etch front being the (100) plane. The {110} planes are etched in KOH about twice as rapidly as {100} planes, while {111} planes are etched at a rate about 100 times slower than for {100} planes [7]

Table 3.2 Liquid, Plasma, and Gas Phase Etchants of Silicon

	$HF:HNO_3:CH_3COOH$	*KOH*	*EDP*	*$N(CH_3)_4OH$ (TMAH)*	*SF_6*	*SF_6/C_4F_8 (DRIE)*	*XeF_2*
Etch type	Wet	Wet	Wet	Wet	Plasma	Plasma	Vapor
Typical formulation	250 ml HF, 500 ml HNO_3, 800 ml CH_3COOH	40 to 50 wt%	750 ml Ethylenediamine, 120g Pyrochatechol, 100 ml water	20 to 25 wt%			Room-temp. vapor pressure
Anisotropic	No	Yes	Yes	Yes	Varies	Yes	No
Temperature	25°C	70°–90°C	115°C	90°C	0°–100°C	20°–80°C	20°C
Etch rate (μm/min)	1 to 20	0.5 to 3	0.75	0.5 to 1.5	0.1 to 0.5	1 to 15	0.1 to 10
{111}/{100} Selectivity	None	100:1	35:1	50:1	None	None	None
Nitride etch (nm/min)	Low	1	0.1	0.1	200	200	12
SiO_2 Etch (nm/min)	10–30	10	0.2	0.1	10	10	0
p^{++} Etch stop	No	Yes	Yes	Yes	No	No	No

(*After:* [3, 6].)

(see Figure 3.6). The latter feature is routinely used to make V-shaped grooves and trenches in (100) silicon wafers, which are precisely delineated by {111} crystallographic planes. The overall reaction consists of the oxidation of silicon followed by a reduction step:

$$Si + 2OH^- \rightarrow Si(OH)_2^{++} + 4e^- \ (oxidation)$$

$$Si(OH)_2^{++} + 4e^- + 4H_2O \rightarrow Si(OH)_6^{--} + 2H_2 \ (reduction)$$

A charge transfer of four electrons occurs during the reaction.

There is little consensus on the origin of the selectivity to {111} crystallographic planes. Proposals made throughout the literature attribute the anisotropy to the lower bond density—and hence lower electron concentration—along {111} planes. Others believe that {111} planes oxidize quickly and are protected during the etch with a thin layer of oxide.

The etch rate of KOH and other alkaline etchants also slows greatly for heavily doped *p*-type (p^{++}) silicon due to the lower concentration of electrons needed for this etch reaction to proceed [7]. P^{++} silicon is thus commonly used as an etch stop. The etch rate of undoped or *n*-type silicon in KOH solutions is approximately 0.5 to 4 μm/min depending on the temperature and the concentration of KOH, but it drops by a factor of over 500 in p^{++} silicon with a dopant concentration above $1 \times 10^{20} cm^{-3}$.

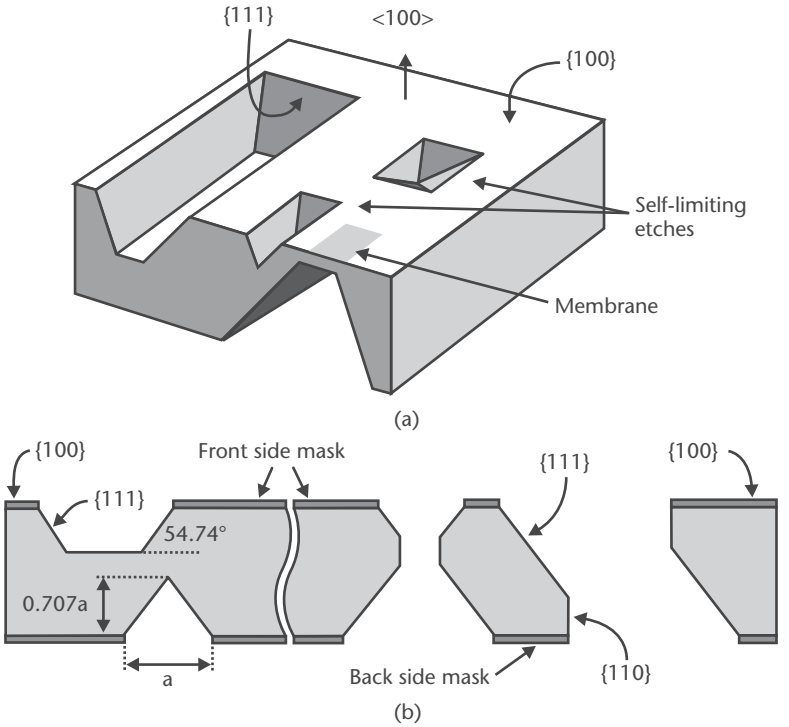


Figure 3.6 Illustration of the anisotropic etching of cavities in {100}-oriented silicon: (a) cavities, self-limiting pyramidal and V-shaped pits, and thin membranes; and (b) etching from both sides of the wafer can yield a multitude of different shapes including hourglass-shaped and oblique holes. When the vertically moving etch fronts from both sides meet, a sharp corner is formed. Lateral etching then occurs, with fast-etching planes such as {110} and {411} being revealed.

LPCVD silicon nitride is an excellent masking material against etching in KOH. Silicon dioxide etches at about 10 nm/min and can be used as a masking layer for very short etches. Photoresist is rapidly etched in hot alkaline solutions and is therefore not suitable for masking these etchants.

Alkali hydroxides are extremely corrosive; aluminum bond pads inadvertently exposed to KOH are quickly damaged. It should be noted that CMOS fabrication facilities are very reluctant to use such etchants or even accept wafers that had previously been exposed to alkali hydroxides for fear of contamination of potassium or sodium, two ions detrimental to the operation of MOS transistors.

In the category of ammonium hydroxides, tetramethyl ammonium hydroxide (TMAH, $N(CH_3)_4OH$) exhibits similar properties to KOH [7]. It etches {111} crystallographic planes 30 to 50 times slower than {100} planes. The etch rate drops by a factor of 40 in heavily *p*-doped silicon ($\sim 1 \times 10^{20} cm^{-3}$). A disadvantage of TMAH is the occasional formation of undesirable pyramidal hillocks at the bottom of the etched cavity. Both silicon dioxide and silicon nitride remain virtually unetched in TMAH and hence can be used as masking layers. It is advisable to remove native silicon dioxide in hydrofluoric acid prior to etching in TMAH because a layer just a few nanometers thick is sufficient to protect the silicon surface from etching. TMAH normally attacks aluminum, but a special formulation containing silicon powder or a pH-controlling additive dissolved in the solution significantly reduces the etch rate of aluminum [8]. This property is useful for the etching of silicon after the complete fabrication of CMOS circuits without resorting to the masking of the aluminum bond pads.

EDP is another wet etchant with selectivity to {111} planes and to heavily *p*-doped silicon. It is hazardous and its vapors are carcinogenic, necessitating the use of completely enclosed reflux condensers. Silicon oxides and nitrides are suitable masking materials for EDP etching. Many metals, including gold, chromium, copper, and tantalum, are also not attacked in EDP; however, the etch rate for aluminum is at about 0.3 μm/min for the formulation given in Table 3.2.

Etching using anisotropic aqueous solutions results in three-dimensional faceted structures formed by intersecting {111} planes with other crystallographic planes. The design of the masking pattern demands a visualization in three dimensions of the etch procession. To that end, etch computer simulation software, such as the program ACES™ available from the University of Illinois at Urbana-Champaign, are useful design tools.

The easiest structures to visualize are V-shaped cavities etched in (100)-oriented wafers. The etch front begins at the opening in the mask and proceeds in the <100> direction, which is the vertical direction in (100)-oriented substrates, creating a cavity with a flat bottom and slanted sides. The sides are {111} planes making a 54.7° angle with respect to the horizontal (100) surface. If left in the etchant long enough, the etch ultimately self-limits on four equivalent but intersecting {111} planes, forming an inverted pyramid or V-shaped trench. Of course, this occurs only if the wafer is thicker than the projected etch depth. Timed etching from one side of the wafer is frequently used to form cavities or thin membranes. Hourglass and oblique-shaped ports are also possible in {100} wafers by etching aligned patterns from both sides of the wafer and allowing the two vertical etch fronts to coalesce and begin etching sideways, then stopping the etch after a predetermined time.

The shape of an etched trench in (110) wafers is radically different (see Figure 3.7). In silicon (110) wafers, four of the eight equivalent {111} planes are perpendicular to the (110) wafer surface. The remaining four {111} planes are slanted at 35.3° with respect to the surface. The four vertical {111} planes intersect to form a parallelogram with an inside angle of 70.5°. A groove etched in (110) wafers has the appearance of a complex polygon delineated by six {111} planes, four vertical and two slanted. Etching in (110) wafers is useful to form trenches with vertical sidewalls, albeit not orthogonal to each other [9].

While concave corners bounded by {111} planes remain intact during the etch, convex corners are immediately attacked (Figure 3.8). This is because any slight erosion of the convex corner exposes fast-etching planes (especially {411} planes) other

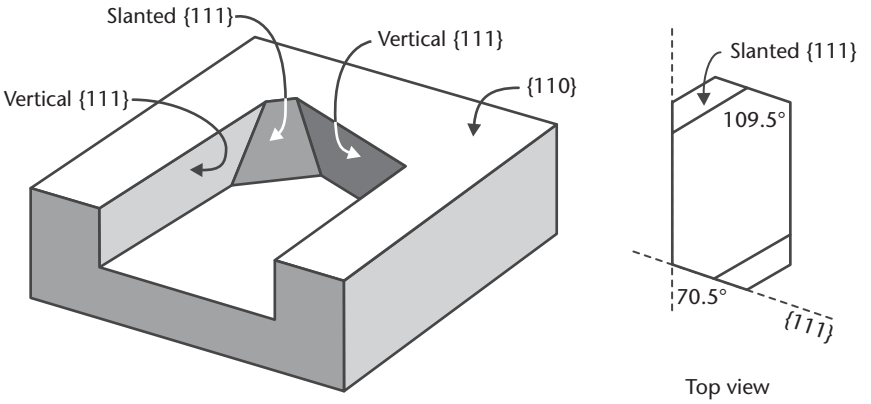


Figure 3.7 Illustration of the anisotropic etching in {110}-oriented silicon. Etched structures are delineated by four vertical {111} planes and two slanted {111} planes. The vertical {111} planes intersect at an angle of 70.5°.

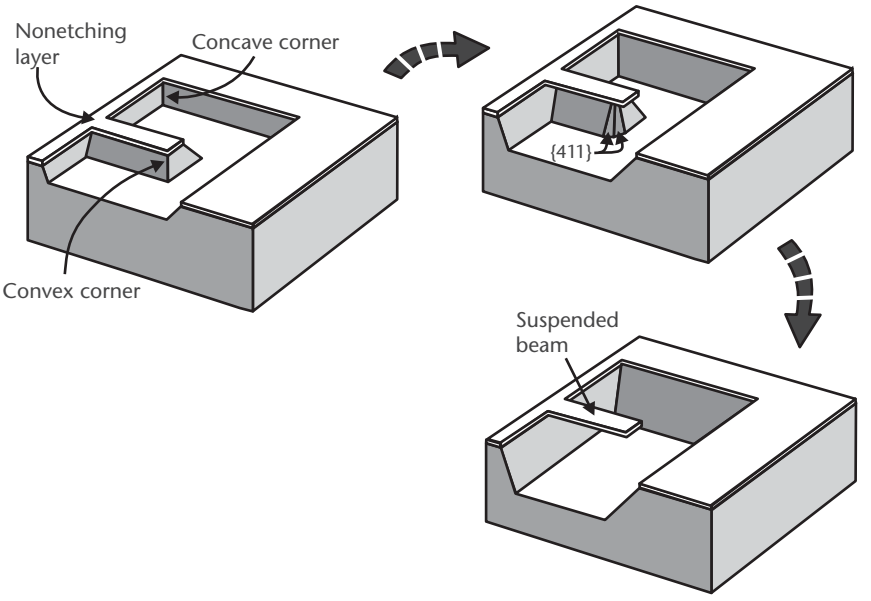


Figure 3.8 Illustration of the etching at convex corners and the formation of suspended beams of a material that is not etched (e.g., silicon nitride, p^{++} silicon). The {411} planes are frequently the fastest etching and appear at convex corners.

than {111} planes, thus accelerating the etch. Consequently, a convex corner in the mask layout will be undercut during the etch; in other words, the etch front will proceed underneath the masking layer. In some instances, such as when a square island is desired, this effect becomes detrimental and is compensated for by clever layout schemes called *corner compensation* [10]. Often, however, the effect is intentionally used to form beams suspended over cavities (see Figure 3.9).

Electrochemical Etching

The relatively large etch rates of anisotropic wet etchants ($>0.5\,\mu$m/min) make it difficult to achieve uniform and controlled etch depths. Some applications, such as bulk-micromachined pressure sensors, demand a thin (5- to 20-μm) silicon membrane with dimensional thickness control and uniformity of better than 0.2 μm, which is very difficult to achieve using timed etching. Instead, the thickness control is obtained by using a precisely grown epitaxial layer and controlling the etch reaction with an externally applied electrical potential. This method is commonly referred as *electrochemical etching* (ECE) [11, 12]. An *n*-type epitaxial layer grown on a *p*-type wafer forms a *p-n* junction diode that allows electrical conduction only if the *p*-type side is at a voltage above the *n*-type; otherwise, no electrical current passes and the diode is said to be in *reverse bias*. During ECE, the applied potential is such that the *p-n* diode is in reverse bias, and the *n*-type epitaxial layer is above its passivation potential—the potential at which a thin passivating silicon dioxide layer forms—hence, it is not etched (see Figure 3.10). The *p*-type substrate is allowed to

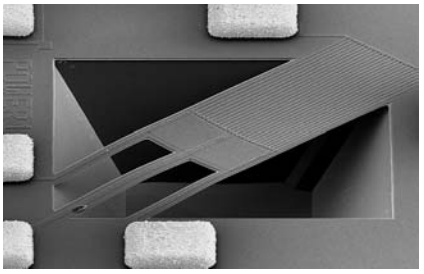


Figure 3.9 Scanning-electron micrograph of a thermally isolated RMS converter consisting of thermopiles on a silicon dioxide membrane. The anisotropic etch undercuts the silicon dioxide mask to form a suspended membrane. (*Courtesy of:* D. Jaeggi, Swiss Federal Institute of Technology of Zurich, Switzerland.)

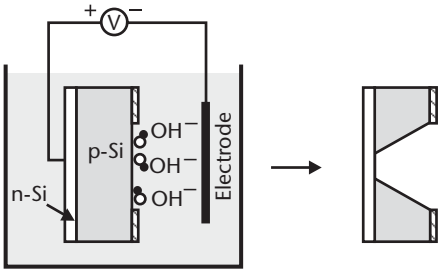


Figure 3.10 Illustration of electrochemical etching using *n*-type epitaxial silicon. The *n*-type silicon is biased above its passivation potential so it is not etched. The *p*-type layer is etched in the solution. The etch stops immediately after the *p*-type layer is completely removed.

electrically float, so it is etched. As soon as the *p*-type substrate is completely removed, the etch reaction comes to a halt at the junction, leaving a layer of *n*-type silicon with precise thickness.

In an original implementation of electrochemical etching on preprocessed CMOS wafers, Reay et al. [13] fabricated a single-crystal *n*-type silicon well with electronic circuits fully suspended from an oxide support beam. Instead of using KOH, they used TMAH with silicon dissolved in the solution in order to prevent the etch of exposed aluminum bond pads (see Figure 3.11).

Plasma Etching

Plasma (or dry) etching is a key process in the semiconductor industry. Companies such as Applied Materials, Inc., of Santa Clara, California, and Lam Research Corp. of Fremont, California, are leading developers and suppliers of plasma-etching systems of silicon as well as silicon dioxide, silicon nitride, and a wide variety of metals. Conventional plasma-phase etch processes are commonly used for etching polysilicon in surface micromachining and for the formation of shallow cavities in bulk micromachining. The introduction in the mid 1990s of deep reactive ion etching (DRIE) systems by Surface Technology Systems (STS), Ltd., of Newport, United Kingdom, Unaxis Semiconductors (formerly PlasmaTherm) of St. Petersburg, Florida, and Alcatel, S.A., of Paris, France, provided a new powerful tool for the etching of very deep trenches (over 500 μm) with nearly vertical sidewalls.

Plasma[2] etching involves the generation of chemically reactive neutrals (e.g., F, Cl), and ions (e.g., SF_x^+) that are accelerated under the effect of an electric field toward a target substrate. The reactive species (neutrals and ions) are formed by the collision of molecules in a reactant gas (e.g., SF_6, CF_4, Cl_2, $CClF_3$, NF_3) with a cloud of energetic electrons excited by an RF electric field. When the etch process is purely chemical, powered by the spontaneous reaction of neutrals with silicon, it is colloquially referred to as *plasma etching*. But if ion bombardment of the silicon surface plays a synergistic role in the chemical etch reaction, the process is then referred to as *reactive ion etching* (RIE). In RIE, ion (e.g., SF_x^+) motion toward the substrate is

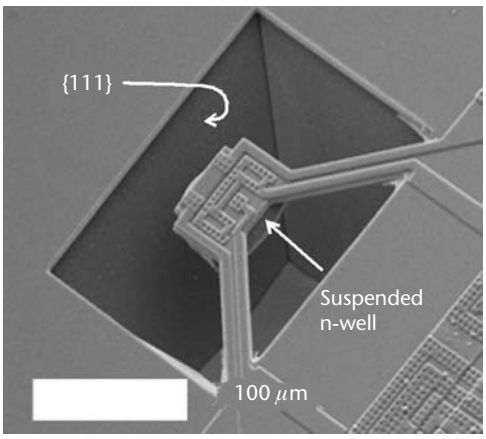


Figure 3.11 A fully suspended *n*-type crystalline silicon island electrochemically etched in TMAH after the completion of the CMOS processing. (*Courtesy of:* R. Reay, Linear Technology, Inc., of Milpitas, California, and E. Klaassen, Intel Corp. of Santa Clara, California.)

nearly vertical, which gives RIE vertical anisotropy. Asymmetric electrodes and low chamber pressures (5 Pa) are characteristic of RIE operation. Inductively coupled plasma reactive ion etching (ICP-RIE) provides greater excitation to the electron cloud by means of an externally applied RF electromagnetic field. Inductively coupled plasma (ICP) increases the density of ions and neutrals resulting in higher etch rates. The ion bombardment energy is controlled by a separate power supply driving the platen on which the wafer sits.

A different, purely physical method of etching is ion milling, in which noble-gas ions (usually argon) are remotely generated, then accelerated at the substrate though a potential on the order of 1 kV. The directionality of the ions results in a very vertical etch profile. Because a chemical reaction is not required, any material can be etched by ion milling. The ion-milling rate is typically much slower than with RIE and varies widely with the material [4].

The remainder of this section focuses on DRIE and its application in micromachining. Further reading on the basics of plasma etching is suggested at the end of this chapter.

DRIE

DRIE evolved in the mid 1990s from the need within the micromachining community for an etch process capable of vertically etching high-aspect-ratio trenches at rates substantially larger than the 0.1 to 0.5 μm/min typical of traditional plasma and RIE etchers. In one approach, developed by Alcatel, the wafer is cooled to cryogenic temperatures. Condensation of the reactant gases (SF_6 and O_2) protects the sidewalls from etching by the reactive fluorine atoms. However, cryogenic cooling may be difficult to maintain locally and could result in undesirable thermal stresses.

Another approach currently used by Alcatel, PlasmaTherm, and Surface Technology Systems (STS) [14] follows a method patented by Robert Bosch GmbH, of Stuttgart, Germany, in which etch and deposition steps alternate in an ICP-RIE system [15] (see Table 3.3). The etch part of the cycle, typically lasting 5 to 15s, uses SF_6, which supplies highly reactive fluorine radicals, to etch silicon. The etch step has both vertical and isotropic character, resulting in a slight mask undercut (see Figure 3.12). In the deposition step, a fluorocarbon polymer (made of a chain of CF_2 groups similar in composition to Teflon™), about 10 nm thick, is plasma-deposited using C_4F_8 as the source gas. In the following etch step, the vertically oriented ions (SF_x^+) enhance

Table 3.3 Process Characteristics of DRIE in the STS System

Etch step	5–15s
SF_6 flow	80–150 sccm
Etch power to coil	600–2,500W
Etch power to platen	5–30W
Deposition step	5–12s
C_4F_8 flow	70–100 sccm
Deposition power to coil	600–1,500W
Pressure	0.5–4 Pa
Platen temperature	0°–20°C
Etch rate	1–15 μm/min
Sidewall angle	90° ± 2°
Selectivity to photoresist	≥ 40 to 1
Selectivity to SiO_2	≥ 100 to 1

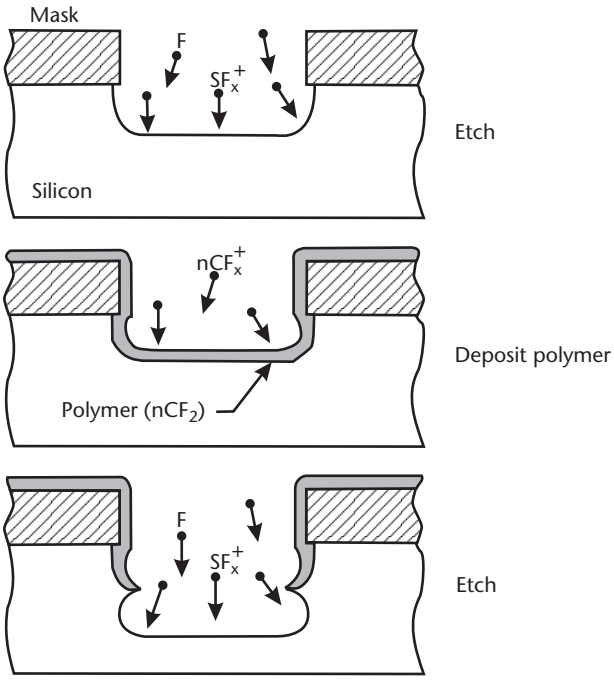


Figure 3.12 Profile of a DRIE trench using the Bosch process. The process cycles between an etch step using SF_6 gas and a polymer deposition step using C_4F_8. The polymer protects the sidewalls from etching by the reactive fluorine radicals. The scalloping effect of the etch is exaggerated.

the effect of fluorine radicals in removing the protective polymer at the bottom of the trench, while the film remains relatively intact along the sidewalls. The repetitive alternation of the etch and passivation steps results in a very directional etch at rates from 1 to over 15 μm/min, depending on the recipe and machine (newer etchers are available with more powerful RF sources). The degree of scalloping—the sidewall texture due to the isotropic component of the etch—varies with the recipe. Recipes optimized for smoother sidewalls can exhibit surface planarity with roughness less than 50 nm, allowing their use as optically reflective surfaces.

A limitation of DRIE is the dependence of the etch rates on the aspect ratio (ratio of height to width) of the trench (see Figures 3.13 and 3.14). The effect is known as lag or aspect-ratio-dependent etching (ARDE). The etch rate is limited by the flux of reactants (namely, F radicals) and drops significantly for narrow trenches. A quick remedy is implemented at the mask layout stage by eliminating large disparities in trench widths. The effect of lag can also be greatly alleviated by

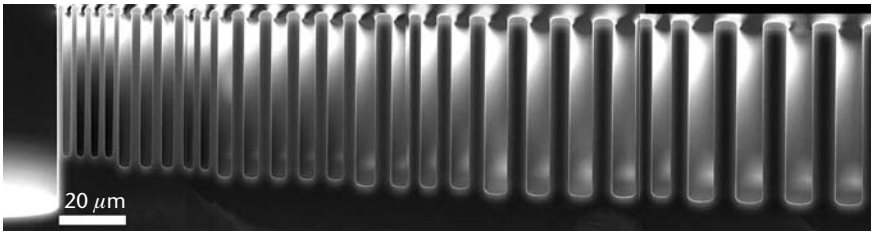


Figure 3.13 ARDE in DRIE. The etch rate decreases with increasing trench aspect ratio. (*Courtesy of:* GE NovaSensor of Fremont, California.)

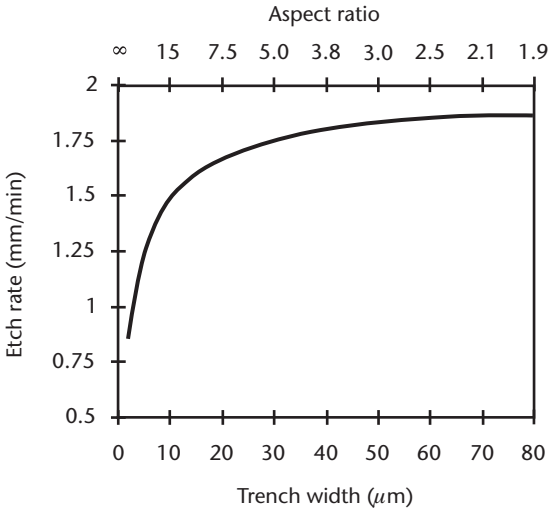


Figure 3.14 Etch-rate dependence on feature size and aspect ratio for a typical DRIE recipe at 600W.

adjusting the process parameters such that a balance is reached between the transport-limited rates of the etch and passivation steps [16]. These parameters are found with experimentation and may vary depending on the mask layout. The penalty for minimizing lag is a reduction in the etch rate to about 1 μm/min.

The high selectivity to silicon dioxide makes it possible to etch deep trenches and stop on a buried layer of silicon dioxide (e.g., silicon-on-insulator wafers). However, when the etch reaches the buried oxide layer, the positive ions charge the oxide, deflecting subsequent ions to the side. The ion bombardment degrades the passivation layer at the bases of the sidewalls, resulting in an undesirable lateral undercut (referred to as *footing* or *notching*) along the silicon-oxide interface (see Figure 3.15). The problem is eliminated in STS DRIE tools by reducing the platen frequency from 13.56 MHz to 380 kHz, which alters the ion energy.

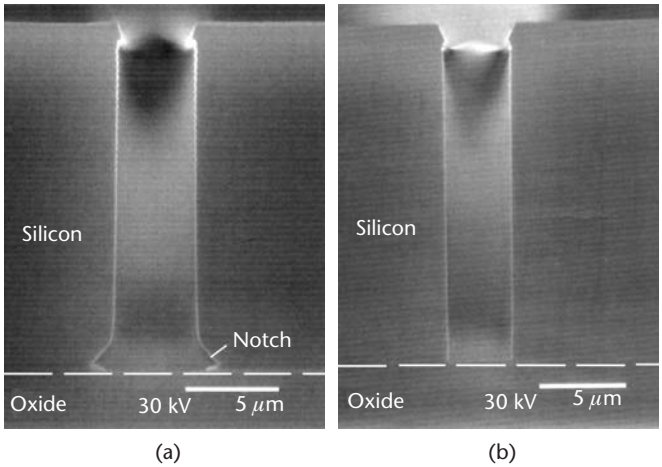


Figure 3.15 (a) Lateral etch observed at the interface between silicon and buried oxide layers, and (b) undercut eliminated with different recipe. (*Courtesy of:* Surface Technology Systems, Ltd., of Newport, United Kingdom.)

DRIE is a powerful tool for the formation of deep trenches with near-vertical sidewalls; however, process development is required for each mask pattern and depth to optimize for low ARDE, good uniformity, high speed, high verticality, small scalloping, and small footing. In general, all cannot be optimized simultaneously. Sequentially running different processes or slowly changing the process as the etch proceeds may be done for the best result.

Advanced Process Tools

Anodic Bonding

Anodic bonding, also known as field-assisted bonding or Mallory® bonding, is a simple process to join together a silicon wafer and a sodium-containing glass substrate (e.g., Corning Pyrex® 7740 and 7070, Schott 8330 and 8329). It is used in the manufacturing of a variety of sensors, including pressure sensors, because it provides a rigid support to the silicon that mechanically isolates it from packaging stress.

The bonding is performed at a temperature between 200° and 500°C in vacuum, air, or in an inert gas environment. The application of 500 to 1,500V across the two substrates, with the glass held at the negative potential, causes mobile positive ions (mostly Na^+) in the glass to migrate away from the silicon-glass interface toward the cathode, leaving behind fixed negative charges in the glass (see Figure 3.16). The bonding is complete when the ion current (measured externally as an electron current) vanishes, indicating that all mobile ions have reached the cathode. The electrostatic attraction between the fixed negative charge in the glass and positive mirror charge induced in the silicon holds the two substrates together and facilitates the chemical bonding of glass to silicon. Silicon dioxide on the silicon surface should be removed before bonding, as a thin (~100 nm) layer is sufficient to disturb the current flow and the bond. A buried oxide layer, such as on a silicon-on-insulator (SOI) wafer, however, does not present a problem, as it conducts sufficiently well at high temperatures to allow the current flow needed for bonding.

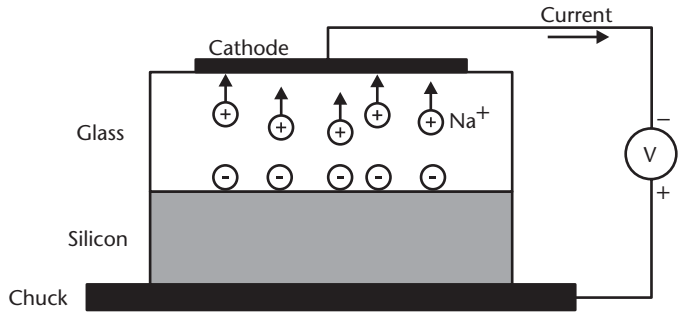


Figure 3.16 Illustration of anodic bonding between glass and silicon. Mobile sodium ions in the glass migrate to the cathode, leaving behind fixed negative charges. A large electric field at the silicon-glass interface holds the two substrates together and facilitates the chemical bonding of glass to silicon.

The coefficient of thermal expansion of the glass substrate is preferably matched to that of silicon in order to minimize thermal stresses. For example, Corning Pyrex® 7740 has a coefficient of thermal expansion of 3.2×10^{-6}/°C; silicon's coefficient is 2.6×10^{-6}/°C at room temperature, rising to 3.8×10^{-6}/°C at 300°C. Sputtered, evaporated, or spin-on glass films containing sodium can also be used to anodically bond two silicon substrates. In this case, the required voltage to initiate the bond process decreases to less than 100V due to the thinness of the glass layer.

Silicon Direct Bonding

Silicon direct bonding, also known as silicon fusion bonding, is a process capable of securely joining two silicon substrates. It emerged as an important step in the development of SOI technology during the 1980s for high-frequency and radiation-hard CMOS applications [17]. SOI wafers made by silicon direct bonding are commercially available today from many vendors. The concept was quickly extended to the manufacture of pressure sensors [18] and accelerometers in the late 1980s and is now an important technique in the MEMS toolbox.

Silicon direct bonding can be performed between two bare single-crystal silicon surfaces or polished polysilicon. One or both surfaces may have thermal or other smooth silicon dioxide or silicon nitride on them. For uniform and void-free bonding, the surfaces must be free of particles and chemical contamination, flat to within about 5 μm across a 100-mm wafer, and smoother than about 0.5- to 1-nm RMS roughness [19] (silicon wafers out of the box are typically on the order of 0.1–0.2 nm RMS roughness).

The direct bonding process starts with cleaning and hydration of the surfaces. The following is a typical sequence, although one or more steps may be swapped or even skipped, as long as the resulting wafers are clean and hydrated. First, the wafers are precleaned in a hot Piranha (sulfuric acid and hydrogen peroxide) solution. Next, they are dipped in a dilute HF solution to etch away the native oxide (or thermal oxide surface) and remove contaminants trapped in the oxide. This is followed by an RCA-1 clean (hot ammonium hydroxide and hydrogen peroxide solution) clean, intended to remove organics. Finally, an RCA-2 clean (hot hydrochloric acid and hydrogen peroxide solution) is done to remove metal contamination. All of the hot hydrogen-peroxide solutions form the hydroxyl (–OH) groups on the surface needed for bonding. This is known as hydration. The bond surfaces are then carefully brought into contact and held together by van der Waals forces [20]. An anneal at 800° to 1,100°C for a few hours promotes and strengthens the bond according to the reaction

$$\mathrm{Si-O-H\bullet\bullet\bullet H-O-Si \rightarrow Si-O-Si + H_2O}$$

In some cases, features on the two bond surfaces must be aligned to each other prior to bonding. For instance, a cavity in one wafer may be joined to an access port provided through the second wafer. Special equipment is necessary to perform the alignment and bonding. SÜSS MicroTec and EV Group, two major equipment manufacturers, use similar schemes to align and bond. The wafers are sequentially mounted in a special fixture and aligned with the two bond surfaces facing each other in a manner similar to double-sided alignment in lithography. A mechanical

clamping fixture holds the aligned wafers in position, separated by thin spacers at the wafer edges. If desired, the fixture and wafers can be placed in a chamber with vacuum, inert gas, oxygen, or other controlled atmosphere. The centers of the wafers are then brought into contact and the spacers removed, allowing the bonded area to proceed from the wafer center to the edge. The relative misalignment is routinely less than 5 μm and can be as good as 1 μm. Direct bonding can be repeated to form thick multiple-wafer stacks, although experience shows that the thicker the stack becomes, the more difficult it is to achieve good bonding [21].

Grinding, Polishing, and Chemical-Mechanical Polishing

Some applications use a thin layer of silicon (5 to 200 μm) that is fusion-bonded to a standard-thickness wafer (525 μm for single-side polished or 400 μm for double-side-polished, 100-mm-diameter wafers), possibly with a layer of oxide between them. Instead of attempting to silicon-fusion bond such a thin, fragile layer to a standard-thickness wafer, two standard-thickness wafers are fusion bonded together, then one side is thinned down to the desired thickness. The thickness reduction is achieved using *grinding* and *polishing*. The wafer stack is mounted on a rotating table and ground by a diamond-bonded wheel spinning in the opposite direction. The grinding mechanically abrades silicon and reduces the thickness of the wafer to near the desired thickness. Hundreds of micrometers can be removed. The resulting surface roughness is removed in the subsequent polishing step in which wafers are mounted inside precise templates on a rotating table. A wheel with a felt-like texture polishes the wafer surface using a slurry containing fine silica or other hard particles in a very dilute alkaline solution (see Figure 3.17). The final surface is smooth, with a thickness control as good as ±0.5 μm. There is frequently invisible damage to the crystal structure incurred during the grinding step that becomes apparent when etched in orientation-dependent etchants. This damage can be removed by growing a thick thermal oxide, then etching it off.

Chemical mechanical polishing, also known as chemical mechanical planarization (CMP), is commonly used in the IC industry for the planarization of dielectric insulating layers. The polishing combines mechanical action with chemical etching using an abrasive slurry dispersed in an alkaline solution (pH > 10). The rate of material removal is controlled by the slurry flow and pH, applied pressure on the polishing head, rotational speed, and operating temperature. CMP is an excellent

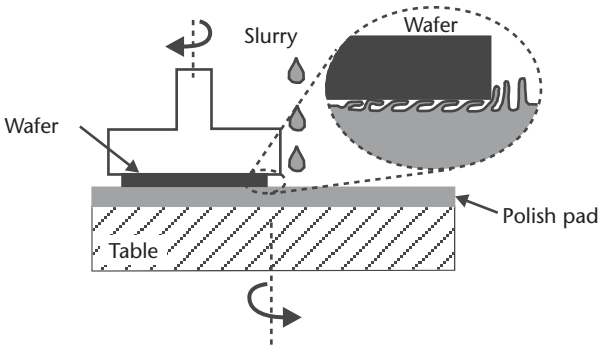


Figure 3.17 Illustration of CMP.

planarization method yielding a surface roughness less than 1 nm over large dimensions, but it is slow with removal rates less than 100 nm/min compared to 1 μm/min for standard polishing.

Sol-Gel Deposition Methods

A *sol-gel* process is a chemical reaction between solid particles in colloidal suspension within a fluid (a sol) to form a gelatinous network (a gel) that can be transformed to solid phase upon removal of the solvent. Sol-gel is not a unique process, but rather represents a broad type of processes capable of forming glasses and ceramics in a multitude of shapes starting from basic chemical precursors. A widespread application of sol-gel processing is in the coating of surfaces with optical absorption or index-graded antireflective materials. It has been used in research laboratories to deposit thick piezoelectric films on silicon substrates.

A sol-gel process starts by dissolving appropriate chemical precursors in a liquid to form a sol (see Figure 3.18). After a time the sol goes through its gel point, the point at which the sol undergoes polymerization, to change it from a viscous liquid state to a gelatinous network. Both sol formation and gelation are low-temperature steps. The gel is then formed into a solid shape (e.g., fiber or lens) or applied as a film coating on a substrate by spinning, dipping, or spraying. For example, TEOS in water can be converted into a silica gel by hydrolysis and condensation using hydrochloric acid as a catalyst. Drying and sintering at an elevated temperature (200°–600°C) results in the transition of the gel to glass and then densification to silicon dioxide [22]. Silicon nitride, alumina, and piezoelectric PZT can also be deposited by sol-gel methods.

Electroplating and Molding

Electroplating is a well-established industrial method that has been adapted in micromachining technology to the patterned deposition of metal films. A variety of metals including gold, copper, nickel, and nickel-iron (Permalloy™) have been electroplated on silicon substrates coated with a suitable thin metal plating base. Table 3.4 lists some plating solutions.

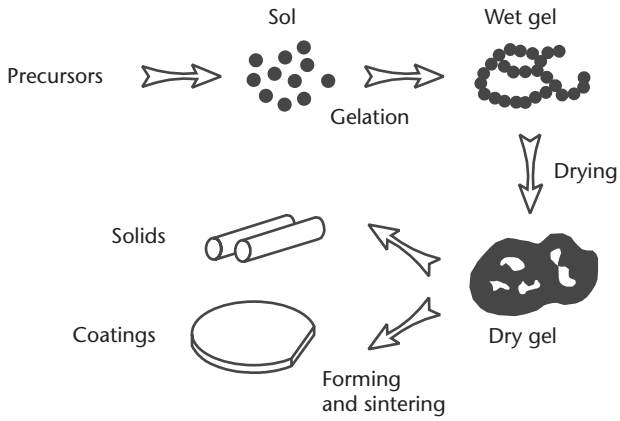


Figure 3.18 Basic flow of a sol-gel process.

Table 3.4 Example Solutions for Electroplating Selected Metals

Metal	*Solution*
Gold	$KAu(CN)_2$:$K_3C_6H_5O_7$:HK_2PO_4:H_2O
	$NaAuSO_3$:H_2O
Copper	$CuSO_4$:H_2SO_4:H_2O
Nickel	$NiSO_4$:$NiCl_2$:H_3BO_3:H_2O
Permalloy	$NiSO_4$:$NiCl_2$:$FeSO_4$:H_3BO_3:$C_7H_4NNaSO_3$:H_2SO_4:H_2O
Platinum	H_2PtCl_6:$Pb(CH_2COOH)_2$:H_2O
Aluminum	$LiAlH_4$:$AlCl_3$ in diethyl ether

Electroplated MEMS structures can take the shape of the underlying substrate and a photoresist mold. First, a conducting seed layer (e.g., of gold or nickel) is deposited on the substrate. In the simplest approach, thick (5- to 100-μm) resist is then deposited and patterned using optical lithography (see Figure 3.19). The largest aspect ratio achievable with optical lithography is approximately three, limited by resolution and depth of focus. In LIGA, optical lithography is replaced with x-ray lithography to define very high aspect ratio features (>100) in very thick (up to 1,000 μm) poly(methylmethacrylate) (PMMA), the material on which Plexiglas® is based. The desired metal is then plated. Finally, the resist and possibly the seed layer outside the plated areas are stripped off.

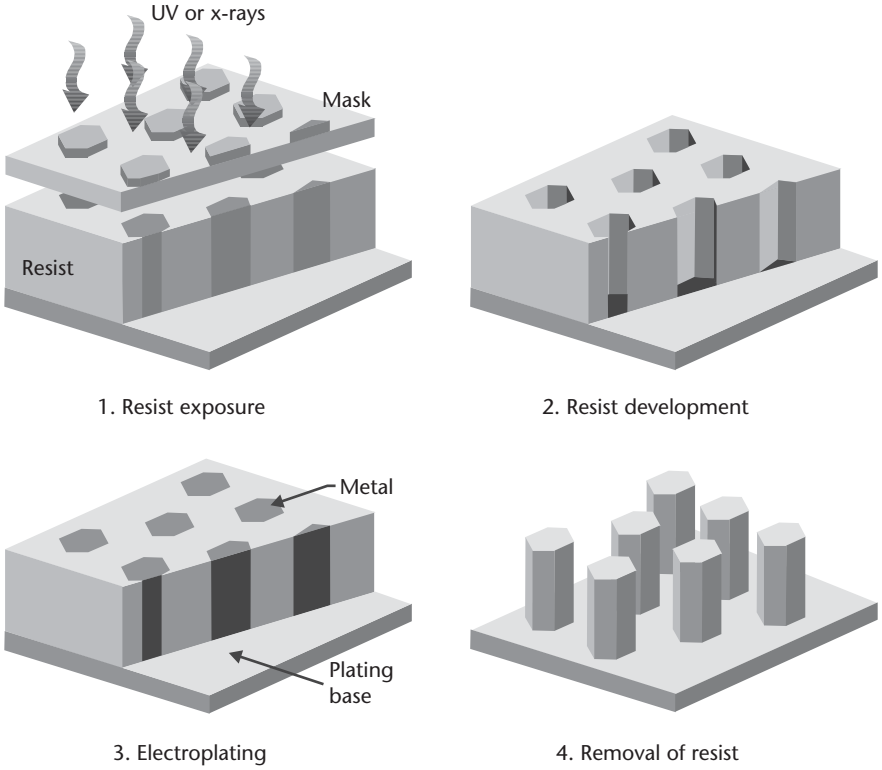


Figure 3.19 Illustration of mold formation using either optical or x-ray lithography and electroplating (LIGA).

The process may be stopped at this point with a metal microstructure suitable for some purposes. Alternatively, the metal can be used as a mold for plastic parts (the "A" in LIGA).

Precision gears and other microstructures have been fabricated using LIGA, but the method is considered expensive because of the requirement to use collimated x-ray irradiation available only from synchrotrons. Mold formation using optical lithography is often called "poor man's LIGA." Guckel [23] provides additional details on the molding of high aspect ratio structures fabricated with x-ray lithography.

In a variation known as electroforming, the plated metal is peeled off of the substrate and *is* the useful structure. Examples of electroformed products are electric shaver screens and some ink-jet heads.

Supercritical Drying

The final step of many micromachining processes is the removal of a sacrificial layer (e.g., using hydrofluoric acid to etch 1 μm of silicon dioxide from under a polysilicon beam). After rinsing, the water must be dried from the wafer. If a freestanding structure overhangs the substrate, surface tension forms a meniscus of water between the two (see Figure 3.20). As the water dries, its volume (and hence thickness) decreases. If the structure is compliant, as is usually the case in surface micromachining, it is pulled down, contacting the substrate. If a sufficiently large, smooth area of the structure makes contact, it can stick, which is known as *stiction* in the micromachining community. Such stuck structures can often be freed by pushing with a probe tip, but this is hardly suitable for production.

A solution to avoid stiction after release is supercritical drying, also known as critical-point drying [24]. In this process, the wafer is moved without drying into methanol, which is miscible with the small amount of water left on the wafer during transfer. The wafer is then placed in a pressure chamber, covered by methanol. Liquid carbon dioxide, which is miscible with methanol, is flowed into the chamber at a pressure of about 7.5–9 MPa as the methanol/carbon dioxide mixture is drained out of the bottom. After a few minutes, only carbon dioxide is left in the chamber. The chamber is then heated from room temperature (near 20°C) to about 35°C, which also increases the pressure (see Figure 3.21). The carbon dioxide has now surpassed the critical point [31.1°C, 7.39 MPa (1071 psia)] and is in the supercritical region, in which liquid and gas are indistinguishable. Finally, the carbon dioxide is vented off. As the pressure drops, the carbon dioxide in the chamber transitions from a supercritical fluid to a gas with only one phase ever being present, thus preventing the

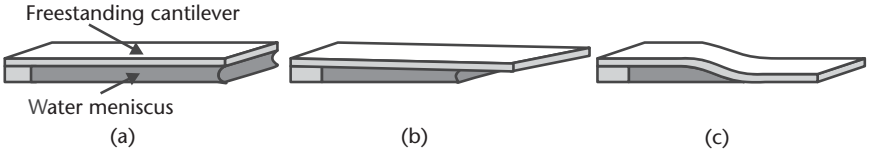


Figure 3.20 Pull-down of a compliant freestanding structure (a cantilever) due to surface tension during drying: (a) water completely fills the volume under the structure; (b) part of the water volume has dried; and (c) most of the water volume has dried, with surface tension pulling the structure down until it touches the substrate.

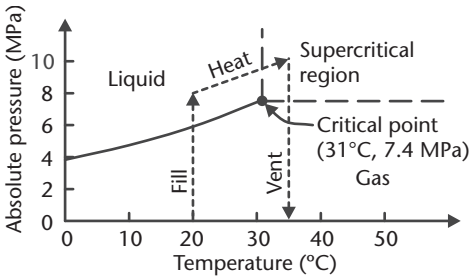


Figure 3.21 The path taken on the carbon dioxide pressure-temperature phase diagram during supercritical drying.

formation of a meniscus and the corresponding stiction. Finally, the dried wafer is removed from the chamber.

Self-Assembled Monolayers

The stiction problem during drying that was presented earlier can also be avoided if a hydrophobic layer is coated onto the structure. One method of doing this is the application of a self-assembled monolayer (SAM) [25]. The SAM precursors used for this application are straight-chain hydrocarbons, such as octadecyltrichlorosilane (OTS, $CH_3(CH_2)_{17}SiCl_3$), with a chemical group at one end that adheres to silicon, silicon-dioxide, and silicon-nitride surfaces. These head groups naturally pack tightly onto the surface and crosslink, leaving the tails sticking straight up away from the surface. The coating self-limits at one molecule of thickness and is hydrophobic.

In a SAM-coating process, the structures are released and rinsed in water as usual, then soaked in a solvent miscible with water. The wafer may be moved to an intermediate solvent compatible with the first solvent and the subsequent SAM solvent. The wafer is then placed in a solution containing the SAM precursor and held for a few minutes, during which the coating occurs. Finally, it is rinsed and dried, which may be done on a hot plate or under a heat lamp. Due to the hydrophobicity of the SAM-coated surface, the contact angle changes, and the water does not pull compliant structures down to the substrate. An added benefit is that if the structure ever does touch down during operation, it will not stick, as it might otherwise do without the coating. SAM coatings have also been studied as a dry lubricant and found to prolong the life of micromachined parts sliding in contact, eventually wearing out [25]. SAMs decompose at high temperatures (~350°C).

SU-8 Photosensitive Epoxy

Epoxies have been in use for decades for joining sections of material together and as a structural component of composites. Some epoxies are formulated to be sensitive to ultraviolet light, allowing photolithographic patterning. SU-8 is a negative-acting photosensitive epoxy intended for use in fabricating microstructures. Originally developed by International Business Machines Corp., it is commercially produced under license by two companies, MicroChem Corp. of Newton, Massachusetts, and SOTEC Microsystems of Renens, Switzerland.

SU-8 is spun onto a substrate in the same manner as photoresist. Different viscosities and a range of spin speeds yield thicknesses from 0.5 to over 250 μm with a single coating [26]. Multiple spins have been used to coat up to 1 mm. The epoxy is then exposed, typically with a standard contact lithography system in the near UV (350–400 nm), but x-rays or an electron beam may also be used. At wavelengths longer than 350 nm, SU-8 has little absorption, allowing exposure through the thickness of much thicker layers than are typically used for traditional photoresist. During exposure, a strong acid is generated where exposed. During the post-exposure bake, the acid initiates thermally driven crosslinking. Immersion in a developer then removes the SU-8 that is not crosslinked. At this point, the remaining material is suitable for many applications, but a hard bake may be performed to promote further crosslinking.

SU-8 structures are the same thickness as the original spin. Aspect ratios (ratio of epoxy height to width) of 20:1 are regularly produced. The cured material is resistant to most chemicals and is thermally stable. SU-8 has been used to form microfluidic channels and optical waveguides. It has also been used as the mask for thick electroplating, although stripping the SU-8 is much more difficult than stripping photoresist.

Photosensitive Glass

Positive-acting photosensitive glass wafers are made commercially under the trade name FORTURAN® by Schott Glas of Mainz, Germany, and processed by Mikroglas Technik AG of Mainz, Germany. FORTURAN is a lithium aluminum silicate glass with small amounts of cerium and silver ions. The full thickness of the glass is exposed with ultraviolet light through a mask, causing the silver ions to form atoms. Annealing causes these atoms to aggregate into microscopic particles, which then serve as nucleation sites for lithium metasilicate crystals. The crystallized volumes are etched relatively rapidly in hydrofluoric acid, leaving holes through the wafer. Up to 14 patterned or unpatterned glass wafers can be thermally bonded together, creating complex systems of channels suitable for microfluidic applications.

Substrates 150 to 1,500 μm thick can be processed. The smallest hole that can be formed in a 400-μm wafer is 60 μm, for an aspect ratio of seven, with a 1.5-μm tolerance. Sidewalls are within 2° of vertical [27].

EFAB

EFAB™ is the trade name for an *electrochemical fabrication* surface micromachining process by Microfabrica, Inc., of Burbank, California, under license from the University of Southern California. In the EFAB process, three-dimensional structures are created by multilayer depositions of patterned metals. Photolithographic techniques are used to deposit a patterned layer of metal (see Figure 3.22). While the details of the process are proprietary, one could accomplish such a structure by electroplating through patterned photoresist. Next, a blanket deposition of a second metal is performed, which fills in the spaces left from the patterned deposition, as well as coating the first metal. The structure is then planarized, leaving the entire substrate covered by patterns of the two metals, all the same thickness. These three steps are then repeated with different masks as many times as necessary to build the desired structure. The definition of each layer is arbitrary with respect to the

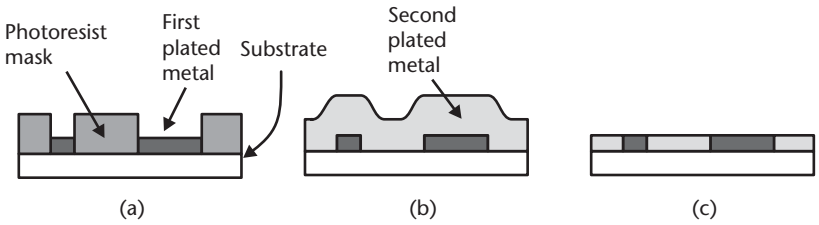


Figure 3.22 The EFAB process: (a) pattern photoresist and selectively electroplate first metal; (b) blanket electroplate second metal; and (c) planarize to same thickness.

previous layer. Finally, one of the metals is selectively etched as a sacrificial layer, leaving behind the other as a structural layer (see Figure 3.23).

Layer thicknesses are in the range of 2 to 20 μm, with a thickness tolerance better than 0.35 μm. Dozens of layers can be formed on 4-in substrates, for an overall stack height of up to several hundred micrometers. The minimum feature size in the plane of the substrate is about 5 μm. One production EFAB process utilizes nickel as the structural material and copper as the sacrificial material. Other material systems to produce copper or nickel-alloy structural layers have been demonstrated.

Nonlithographic Microfabrication Technologies

Several conventional, non-IC-related technologies that do not use photolithography are also capable of forming features of relatively small dimensions. These include mechanical machining, ultrasonic machining, electrodischarge machining, and laser machining. Only some of these can be considered to be batch fabrication. As these fabrication methods have been in use for decades, they have had time to evolve,

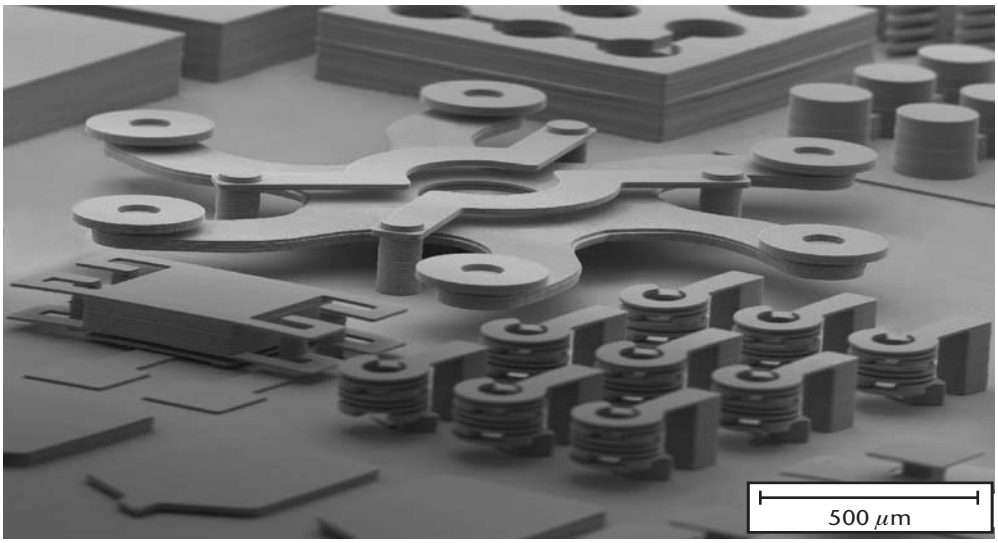


Figure 3.23 EFAB example demonstrating the complex three-dimensional structures that can be produced. The layers of metal are clearly visible. (*Courtesy of:* Microfabrica Inc., of Burbank, California.)

yielding ever lower cost and finer dimensional control. In some applications, such as ink-jet printer nozzles and automobile fuel-injection nozzles, photolithographic fabrication methods have been used, but proved less economical than the more established methods. In addition to competing with lithographic technologies, non-IC-related fabrication technologies are often used in conjunction with them in the production of a final product; examples include bulk-micromachined pressure sensors with ultrasonically drilled glass bonded to the back side and ink-jet heads with surface-micromachined heaters and laser-drilled ports. Two newer techniques for creating submicrometer patterns are also discussed in this section.

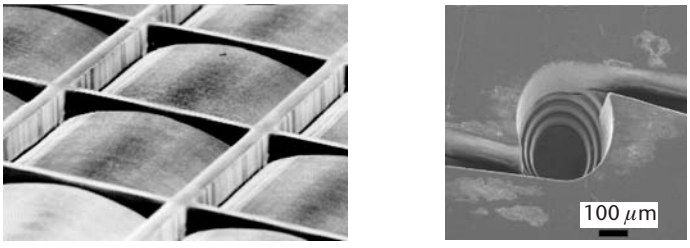
Ultraprecision Mechanical Machining

Cutting tools such as mills, lathes, and drills using a specially hardened cutting edge have been in use for the production of macroscopic parts for over a century. Using modern computer-numerical-controlled (CNC) machines with sharply tipped diamond-cutting tools, many metals and even silicon have been milled to a desired shape, with some features smaller than 10 μm. Many of these shapes, such as retrograde undercuts with flat sidewalls, cannot be formed using lithographic methods. Resolution of about 0.5 μm can be achieved, with surface roughnesses on the order of 10 nm [28]. Example applications include optical mirrors and computer hard-drive disks.

Laser Machining

Focused pulses of radiation, typically 0.1–100 ns in duration, from a high-power laser can ablate material (explosively remove it as fine particles and vapor) from a substrate. Incorporating such a laser in a CNC system enables precision laser machining. Metals, ceramics, silicon, and plastics can be laser machined. Holes as small as tens of microns in diameter, with aspect ratios greater than 10:1, can be produced. Arbitrary shapes of varying depths are laser machined by scanning the beam to remove a shallow layer of material, then scanning again until the desired depth has be reached (see Figure 3.24). Laser machining can be used to create perforations in silicon wafers for subsequent cleaving to form individual chips, as well as simply cutting though the full wafer thickness.

Laser machining is most often a serial process, but with mask-projection techniques, it becomes a parallel process. It has successfully competed with KOH etching



(a) (b)

Figure 3.24 Laser machining examples: (a) microlenses in polycarbonate; and (b) fluid-flow device in plastic. Multiple depths of material can be removed. (*Courtesy of:* Exitech Ltd., of Oxford, United Kingdom.)

and with electroplating in the production of ink-jet nozzles. Due to its speed, low cost, and rapid turn-around time, laser machining is one of the preferred methods of creating trenches and cuts in plastics.

Electrodischarge Machining

Electrodischarge machining, also called electrical-discharge machining or sparkerosion machining (EDM) uses a series of electrical discharges (sparks) to erode material from a conductive workpiece. High-voltage pulses, repeated at 50 kHz to 500 kHz, are applied to a conductive electrode, typically made of graphite, brass, copper, or tungsten. Electrodes as small as 40 μm in diameter have been used, limiting features to about the same size. Features with aspect ratios of over 10 can be fabricated, with a surface roughness on the order of 100 nm. Each discharge removes a small volume of material, typically in the range of 10^3 to 10^5 μm^3, from the workpiece [29]. EDM is performed in a dielectric liquid such as mineral oil. Due to heating, a gas bubble is formed during each voltage pulse. After the pulse, the bubble collapses, flushing away debris from the blank and electrode.

EDM has been used to create the tooling for molds and stamping tools, as well as final products such as nozzles and holes in microneedles.

Screen Printing

Screen printing, also known as silk screening, has been used for the printing of images for millennia. In electronics, it has long been used in the production of ceramic packages and more recently for large flat-panel displays. In a parallel process, many ceramic packages are processed together on a single plate, then separated near the end of the process. A wide variety of materials, including metals and ceramics, can be applied using screen printing. It does not have same resolution as photolithography, but is cost effective and is readily applied to large substrates.

Screen printing begins with the production of a stencil, which is a flat, flexible plate with solid and open areas (see Figure 3.25). The stencil often has a fine-mesh screen as a bottom layer to provide mechanical rigidity. Separately, a paste is made of fine particles of the material of interest, along with an organic binder and a solvent. A mass of paste is applied to the stencil, then smeared along with a squeegee. A thin layer of paste is forced though the openings in the stencil, leaving a

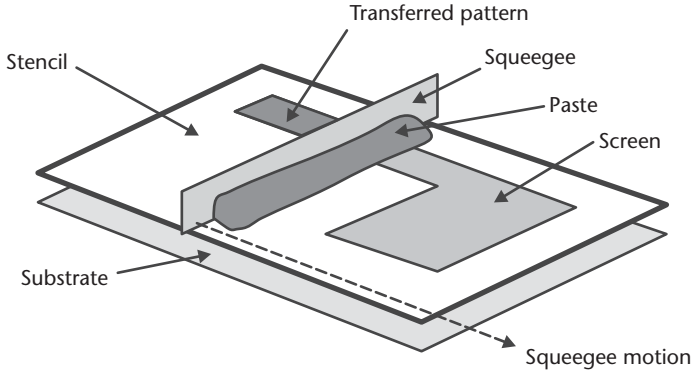


Figure 3.25 Illustration of screen printing.

pattern on the underlying substrate. Drying evaporates the solvent. Firing burns off the organic binder and sinters the remaining metal or ceramic into a solid, resulting in a known amount of shrinkage. Metal lines with 125-μm lines and spaces are made in the production of ceramic packaging (discussed further in Chapter 8) [30], with 30-μm features demonstrated [31]. Film thicknesses after firing range from roughly 10 to 200 μm. Multiple layers of different materials can be stacked.

Microcontact Printing/Soft Lithography

Microcontact printing, a microscale form of ink printing also called *soft lithography,* has been studied by several research groups [32, 33]. It enables low-cost production of submicrometer patterns and has been studied as an alternative to conventional photolithography, but is not presently a product fabrication method.

The process begins with the production of the original, hard, three-dimensional master pattern (see Figure 3.26), which can involve conventional photolithography and etching, electron-beam lithography, laser scribing, diamond scribing, or any other suitable method. A mold of an elastomer, usually poly(dimethylsiloxane) (PDMS), is made against the master, then peeled off to create a stamp with raised patterns. An "ink," a liquid solution typically of an alkanethiol (a hydrocarbon chain ending in a thiol, an –SH group) such as hexadecanethiol, is poured onto the PDMS stamp and dried. The inked stamp is then held against a substrate coated with gold, silver, or copper, then removed. The thiol end of each "ink" molecule bonds to the metal, forming a densely packed, single-molecule-thick coating of hexadecanethiol where the raised areas of the stamp were. Such SAM coatings can be envisioned as similar to turf with dense blades of grass. Once the SAM coating is in place, it can be used as an etch mask for the metal. The metal can then be used as an etch mask for the underlying substrate, such as silicon.

Several variations on this scheme may be performed. In one, a metal catalyst "ink" is stamped on the substrate, which is then used for the selective plating of copper. In another, proteins or other biological molecules are coated onto a flat stamp. A patterned PDMS layer contacts the flat stamp and is removed, taking the protein

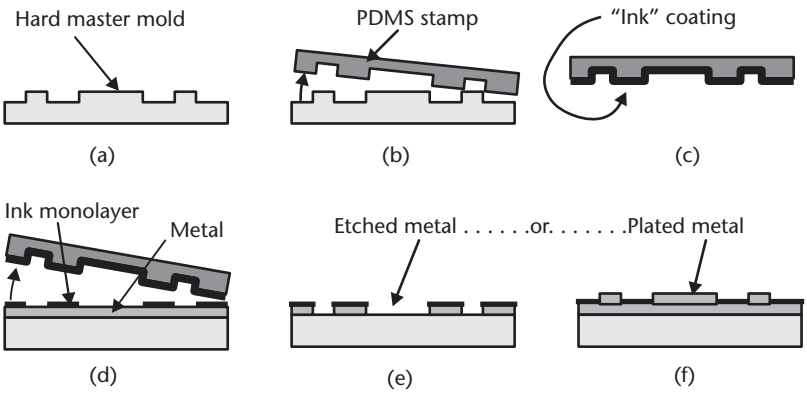


Figure 3.26 Microcontact printing: (a) create master; (b) form PDMS stamp and peel off; (c) coat with "ink"; (d) press inked stamp against metal and remove, leaving ink monolayer; (e) use self-assembled monolayer as an etch mask; or (f) as a plating mask.

with it where contact occurred. The flat stamp is then held against a substrate, transferring the protein pattern [33].

Features smaller than 0.1 μm have been made using microcontact printing. The best alignment accuracy of a second pattern, however, is at present about 20 μm [33], so most soft lithography applications have used a single step.

Nanoimprint Lithography

As with microcontact printing, nanoimprint lithography has the goal of generating submicrometer features at low cost and high throughput and is not a production process [34]. It starts with a mold of etched silicon, silicon dioxide, or other hard material created using optical or electron-beam lithography (see Figure 3.27). Separately, a substrate is coated with a 50- to 250-nm resist layer such as PMMA or a more conventional novolak-resin-based resist, which does not need to be photosensitive. The resist is heated above its glass transition temperature so that it flows easily under pressure. The mold is then pressed into the resist, which flows to the sides of the high points in the mold. The mold is removed, leaving an unintentional residue of resist where the mold high points were. This residue is stripped using vertical RIE. At this point, the resist pattern can be used like conventional photoresist in an etch, liftoff, or plating process.

Features 25 nm wide with smooth sidewalls have been demonstrated. Alignment accuracy of a second nanoimprint step is likely to be many micrometers, but the technique has been combined with optical lithography to fabricate devices with several layers.

Hot Embossing

In the hot embossing process, a pattern in a master is transferred to a thermoplastic material. If the dimensions are relatively large (>100 μm), the master can be made with conventional machining. Smaller dimensions can be produced using nickel electroplated through patterned photoresist. The master is pressed into the thermoplastic (e.g., PMMA, polycarbonate, polypropylene) just above the material's glass transition temperature. The master and plastic are cooled while in contact, then separated, leaving a pattern in the plastic.

Hot embossing is used in microfluidics for creating trenches in substrates of thermoplastic. Several substrates can then be bonded together to form channels for a microfluidic system. Aspect ratios over 10 can be achieved, with the minimum feature size limited by the master.

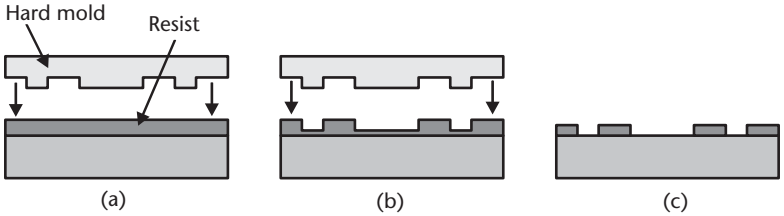


Figure 3.27 Nanoimprint lithography: (a) press hard mold into resist coating; (b) remove mold; and (c) RIE to remove residue (*After:* [34].)

Ultrasonic Machining

In ultrasonic machining, also known as ultrasonic impact grinding, a transducer vibrates a tool at high frequency (20–100 kHz). The tip of the tool is pushed against the workpiece as a slurry of water or oil and abrasive particles, such as boron carbide, aluminum oxide, or silicon carbide, is flushed across the surface. There are several mechanisms for removal of material: The tool vibration directly hammers particles into the surface, as well as imparting a high velocity to other particles, both of which chip away at the workpiece. Cavitation erosion and chemical action can also contribute. The microscopic chips are carried away by the slurry. As the tool moves slowly into the workpiece, a hole with vertical sidewalls is created. An array of tips can drill many holes at the same time; Figure 3.28 shows examples in several materials. The hole shape matches that of the tool and can be round, square, or other.

Ultrasonic machining can be performed on hard, brittle materials (with a Knoop hardness above about 400) such as glasses, ceramics, diamond, and silicon. The minimum hole diameter is about 150 μm. At the other extreme, holes over 100 mm have been machined. For small holes, the maximum aspect ratio is about five, increasing to over 15 for holes several millimeters in diameter. With tolerancing, the size accuracy of 1-mm holes is typically ±50 μm, improving to ±25 μm for larger holes. Hole depth can be over 10 mm.

Combining the Tools—Examples of Commercial Processes

The sequence in which various processes from the toolbox are combined determines a unique microfabrication process. The process may be specific to a particular design or may be sufficiently general that it can be used to fabricate a range of designs. This section describes four example fabrication processes that are generic in their nature and used today in manufacturing at a number of companies and commercial foundries.

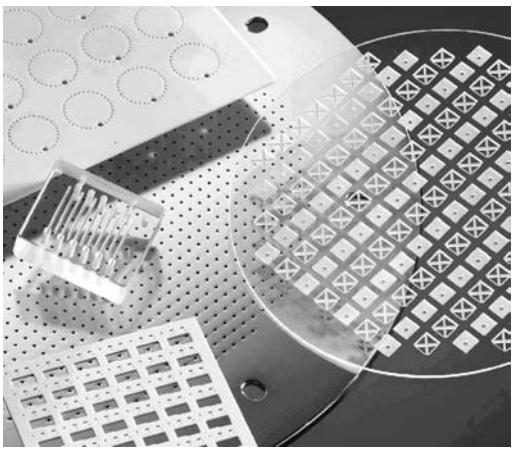


Figure 3.28 Photograph of ultrasonically drilled holes and cavities in glass (clear), alumina ceramic (white), and silicon (shiny). All of the holes in a single substrate are drilled simultaneously. (*Courtesy of:* Bullen Ultrasonics, Inc., of Eaton, Ohio.)

All of these processes are compatible with CMOS fabrication and hence allow the integration of electronic circuits alongside microelectromechanical devices. Successful integration requires that circuit and structural processing steps do not adversely affect each other; for example, once aluminum is on the wafer in contact with silicon, it cannot be heated above 400–450°C. As will be observed, a key distinguishing feature among the processes is the *release* step that frees the microstructures in selected locations from the underlying substrate.

Polysilicon Surface Micromachining

In surface micromachining, thin layers of a material—most commonly polysilicon—form the structural elements. Originating at the University of California at Berkeley, polysilicon surface micromachining is an established manufacturing process at Analog Devices, Inc., of Norwood, Massachusetts, MEMSCAP (formerly Cronos Integrated Microsystems, Inc.) of Research Triangle Park, North Carolina, and Robert Bosch GmbH of Stuttgart, Germany. Bustillo et al. present a comprehensive review of surface micromachining in a special issue of the *Proceedings of the IEEE* on MEMS [35].

Polysilicon surface micromachining combines a stack of patterned polysilicon thin films with alternating patterned layers of sacrificial silicon dioxide. A single layer of structural polysilicon is sufficient to make many useful devices, and up to five polysilicon and five oxide layers are a standard process at Sandia National Laboratories of Albuquerque, New Mexico. The polysilicon is deposited using LPCVD, followed by a high-temperature anneal (>900°C) to relieve mechanical stress. The silicon dioxide is deposited using LPCVD or PECVD and is often doped with phosphorus [phosphosilicate glass (PSG)] to increase the etch rate in hydrofluoric acid. In the Sandia process, the polysilicon and silicon dioxide layers are each 2 μm thick. By contrast, Robert Bosch uses a process with 10-μm-thick polysilicon grown by epitaxy over silicon dioxide.

Each of the layers in the stack is lithographically patterned and etched before the next layer is deposited in order to form the appropriate shapes and to make provisions for anchor points to the substrate (see Figure 3.29). The final release step consists of etching the silicon dioxide (hence the sacrificial term) in a hydrofluoric acid solution to free the polysilicon plates and beams, thus allowing motion in the plane of and perpendicular to the substrate. Small holes are usually added to large plates to allow the sacrificial etchant access for faster release. To avoid sticking of compliant structures when drying the wafer, supercritical drying or a self-assembled monolayer is often used.

Gears, micromotors, beams, simple as well as hinged plates, and a number of other structures have been demonstrated, though primarily accelerometers and yaw-rate sensors are currently in high-volume production. Surface micromachining offers significant flexibility to fabricate planar structures one layer at a time, but their thinness limits the applications to those benefiting from essentially two-dimensional forms.

Polysilicon is a useful structural material because integrated circuit processes already exist for depositing and etching it and because its thermal coefficient of expansion is well matched to that of the silicon substrate. However, surface micromachining is not limited to the materials just described. Many systems of

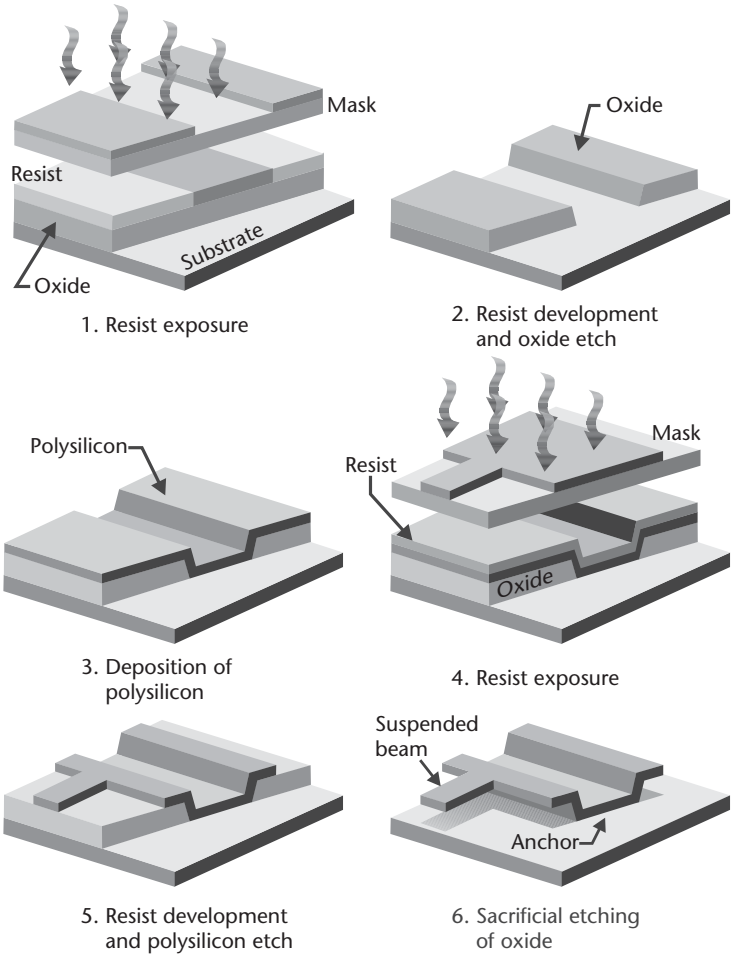


Figure 3.29 Schematic illustration of the basic process steps in surface micromachining.

structural layer, sacrificial layer, and etchant have been used, as shown in Table 3.5. The etchant must etch the sacrificial layer at a useful rate, while having little or no impact on the structural layer. Reasons for selecting materials other than polysilicon include the need for higher electrical conductivity, higher optical reflectivity, and lower deposition temperature for compatibility with CMOS circuitry that is already on the wafer. For example, Texas Instruments' Digital Mirror Device™ (DMD™) display technology uses a surface-micromachined device with aluminum as its

Table 3.5 Some Systems of Materials for Surface Micromachining

Structural Material	*Sacrificial Material*	*Etchant*
Polysilicon	Silicon dioxide/PSG	Hydrofluoric acid
Silicon nitride	Silicon dioxide/PSG	Hydrofluoric acid
Silicon nitride	Polysilicon	Potassium hydroxide; xenon difluoride
Gold, tungsten, molybdenum, other metals	Silicon dioxide/PSG	Hydrofluoric acid
Aluminum	Photoresist/organic	Oxygen plasma
Nickel	Copper	Ammonium persulfate
Silicon-germanium	Germanium	Hydrogen peroxide
Silicon carbide	Silicon dioxide	Hydrofluoric acid

structural element and an organic polymer as a sacrificial layer. Chapter 5 describes this particular device in greater detail.

Combining Silicon Fusion Bonding with Reactive Ion Etching

The silicon fusion bonding with reactive ion etching (SFB-DRIE) process involves the formation of tall structures in crystalline silicon to overcome the thinness limitation of surface micromachining [36]. Instead of depositing thin polysilicon layers, crystalline silicon substrates are fusion bonded to each other in a stack. Each substrate is polished down to a desired thickness, then patterned and etched before the next one is bonded. An optional intermediate silicon dioxide between the silicon substrates is not a sacrificial layer but is rather for electrical and thermal insulation. The process allows the building of complex three-dimensional structures one thick layer at a time.

The basic process flow begins by etching a cavity in a first wafer, referred to as the handle wafer (see Figure 3.30). A second wafer is silicon fusion bonded on. An optional grind and polish step reduces the thickness of the bonded wafer to any desired value. CMOS electronic circuits can then be integrated on the top surface of the bonded stack without affecting any of its mechanical properties. Finally, a DRIE step determines the shape of the microstructures and mechanically releases them as soon as the etch reaches the embedded cavity. This cavity takes the role of the sacrificial layer in surface micromachining and ensures that the micromechanical structures are free to move except at well-defined anchor points.

The high aspect ratio and depth available using the SFB-DRIE process add new dimensions to the design and fabrication of complex three-dimensional structures (see Figure 3.31). A range of new applications, including those integrating fluid flow functions such as valving and pumping, can be addressed with this process. Robust thermal actuators made of crystalline silicon are also feasible with an available output force approaching one newton. This process is now a manufacturing platform at GE NovaSensor of Fremont, California.

DRIE of SOI Wafers

The availability of double-sided aligners, DRIE tools, and SOI wafers led to a relatively simple process for fabricating three-dimensional microstructures that became popular in the late 1990s. The process begins with DRIE of the thinner top layer of an SOI wafer to form the desired structure (see Figure 3.32). The etch stops with high selectivity on the buried oxide layer. If undercut of the silicon at the oxide interface control is not desired, the specialized stop-on-oxide recipe discussed earlier can be used. A large area of the back side, corresponding to the structure on the front side, is etched to the buried oxide layer. Finally, the now-freestanding buried oxide is etched away, typically with hydrogen fluoride [hydrofluoric acid (HF)] vapor or a liquid HF solution, both of which selectively etch the oxide. If liquid HF is used and the structure is fragile, it must be handled carefully to avoid breakage during etching, rinsing, and drying. A variation on the process is to etch the device structure from the top, then release it by etching the underlying oxide, which may be as thick as 2 μm, in liquid HF. If the structure is sufficiently stiff, it can be dried without special handling. If it is too compliant, critical-point drying can be used. Similar processes are in development or commercial use by companies including the

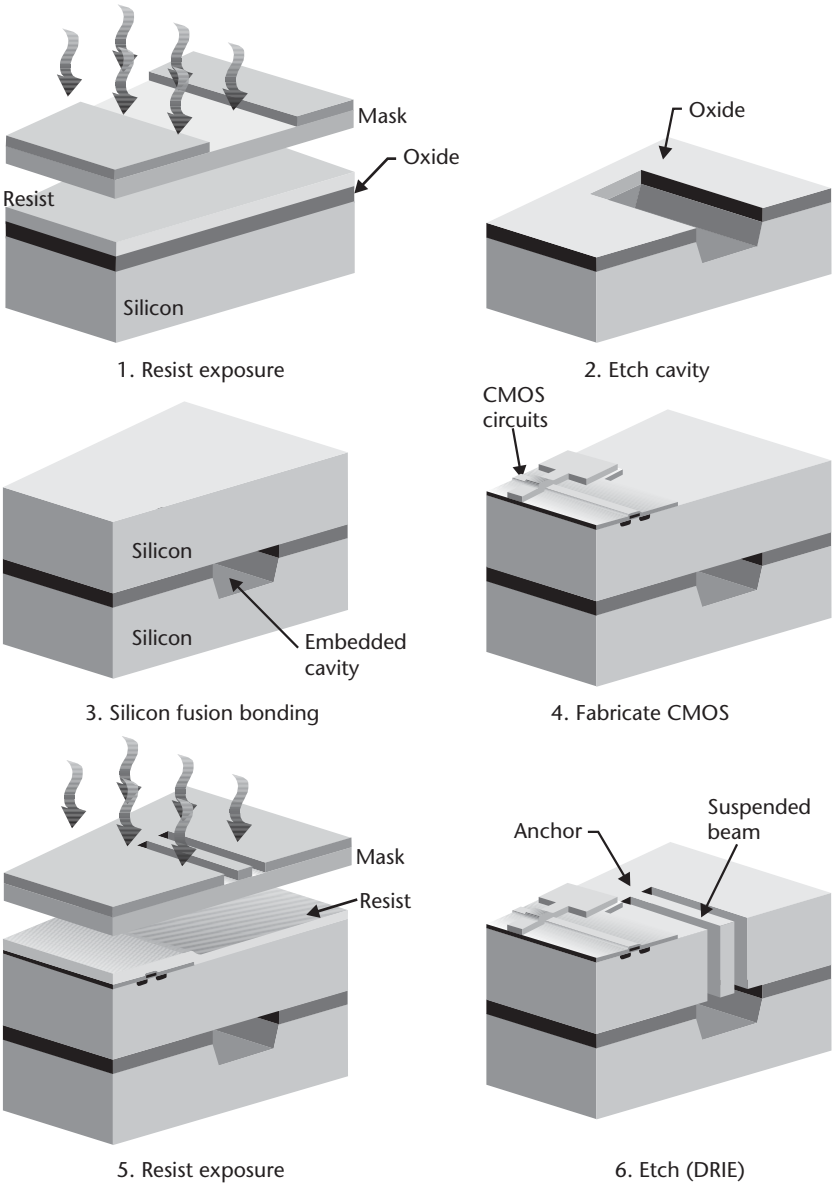


Figure 3.30 Fabrication process combining silicon fusion bonding and DRIE.

Micromachined Products Division of Analog Devices, Inc., of Belfast, United Kingdom, TRONIC'S Microsystems SA of Grenoble, France, and DiCon Fiberoptics, Inc., of Richmond, California (see Figure 3.33).

Single Crystal Reactive Etching and Metallization

The single-crystal reactive etching and metallization (SCREAM) process [37] uses yet another approach to release crystalline microstructures. Standard lithography and etching methods define trenches between 10 and 50 μm in depth, which are then coated on the top, sidewalls, and bottom with a conformal layer of PECVD silicon dioxide (see Figure 3.34). An anisotropic etch step selectively removes the protective

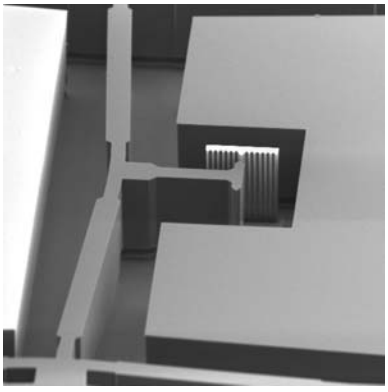


Figure 3.31 Scanning electron microscope image of a 200-μm-deep thermal actuator fabricated using silicon fusion bonding and DRIE. (*Courtesy of:* GE NovaSensor of Fremont, California.)

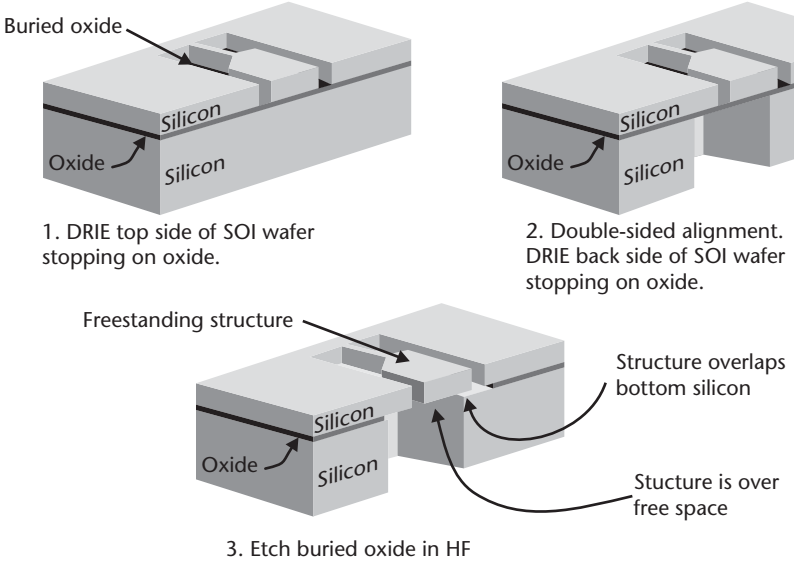


Figure 3.32 Example process for DRIE of SOI wafers. The final structure may be over free space or can overlap the bottom wafer (or both, as in this example).

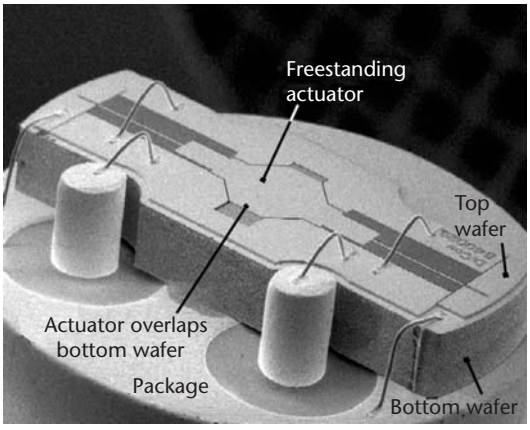


Figure 3.33 Scanning electron microscope image of a variable optical attenuator made by DRIE of an SOI wafer. (*Courtesy of:* DiCon Fiberoptics, Inc., of Richmond, California.)

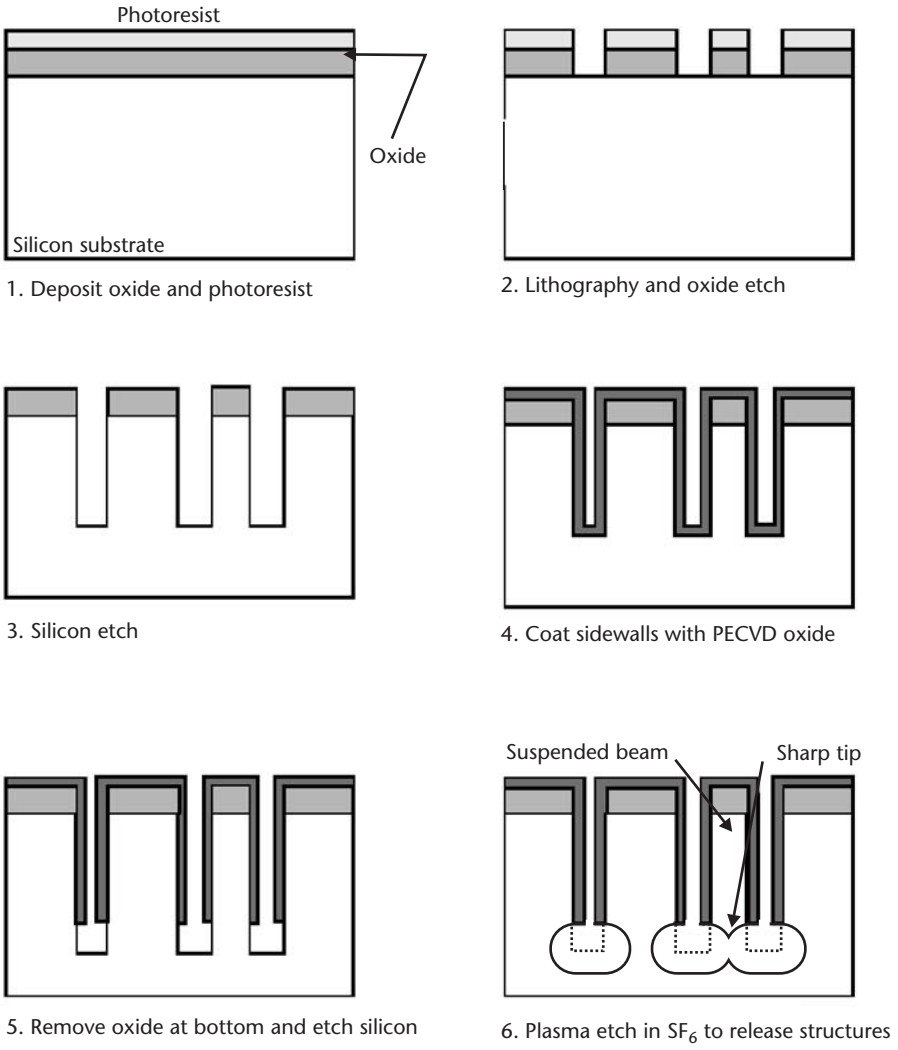


Figure 3.34 Basic steps of the SCREAM process. (*After:* [37].)

oxide only at the bottom of the trench. A subsequent plasma silicon etch extends the depth of the trench. A dry isotropic etch step using sulfur hexafluoride (SF_6) laterally etches the exposed sidewalls near the bottom of the trench, thus undercutting adjacent structures and mechanically releasing them. Sputter deposition of aluminum provides the metal for electrical contacts and interconnects.

This process, known by its SCREAM acronym, was initially developed at Cornell University. Kionix, Inc., of Ithaca, New York, uses a variation of SCREAM for the manufacture of accelerometers, micromirrors, and other devices.

Summary

The toolbox of micromachining processes is very large and diverse. The vast majority of the methods can be condensed into three major categories:

- Material deposition, including thin film deposition and bonding processes;
- Pattern definition using lithography;
- Etching and mechanical material removal.

A complete micromachining process flow consists of a series of steps using a number of methods from the toolbox to build complex microstructures one layer at a time.

References

[1] Katz, L. E., "Oxidation," in *VLSI Technology*, S. M. Sze (ed.), New York: McGraw-Hill, 1983, pp. 131–167.

[2] Thornton, J. A., and D. W. Hoffman, "Stress Related Effects in Thin Films," *Thin Solid Films*, Vol. 171, 1989, pp. 5–31.

[3] Williams, K. R., and R. S. Muller, "Etch Rates for Micromachining Processing," *Journal of Microelectromechanical Systems*, Vol. 5, No. 4, December 1996, pp. 256–269.

[4] Williams, K. R., K. Gupta, and M. Wasilik, "Etch Rates for Micromachining Processing–Part II," *Journal of Microelectromechanical Systems*, Vol. 12, No. 6, December 2003, pp. 761–778.

[5] Williams, K., "Etching," in *Properties of Crystalline Silicon*, R. Hull (ed.), London: INSPEC, The Institution of Electrical Engineers, 1999, Chapter 16.

[6] Kovacs, G. T. A., N. I. Maluf, and K. E. Petersen, "Bulk Micromachining of Silicon," *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1536–1551.

[7] Seidel, H., et al., "Anisotropic Etching of Crystalline Silicon in Alkaline Solutions," *Journal of Electrochemical Society*, Vol. 137, No. 11, November 1990, pp. 3612–3632.

[8] Schnakenberg, U., W. Benecke, and P. Lange, "TMAHW Etchants for Silicon Micromachining," *Proc. 1991 Int. Conf. on Solid-State Sensors and Actuators*, San Francisco, CA, June 24–27, 1991, pp. 815–818.

[9] Ammar, E. S., and T. J. Rodgers, "UMOS Transistors on (110) Silicon," *IEEE Transactions on Electron Devices*, Vol. ED-27, No. 5, May 1980, pp. 907–914.

[10] Sandmaier, H., et al., "Compensation Techniques in Anisotropic Etching of (100)-Silicon Using Aqueous KOH," *Proc. 1991 Int. Conf. on Solid-State Sensors and Actuators*, San Francisco, CA, June 24–27, 1991, pp. 456–459.

[11] Waggener, H. A., "Electrochemically Controlled Thinning of Silicon," *Bell System Technology Journal*, Vol. 50, 1970, pp. 473–475.

[12] Kloeck, B., et al., "Study of Electrochemical Etch-Stop for High Precision Thickness Control of Silicon Membranes," *IEEE Transactions on Electron Devices,* Vol. 36, No. 4, April 1989, pp. 663–669.

[13] Reay, R. J., E. H. Klaassen, and G. T. A. Kovacs, "Thermally and Electrically Isolated Single-Crystal Silicon Structures in CMOS Technology," *IEEE Electron Device Letters*, Vol. 15, October 1994, pp. 309–401.

[14] Bhardwaj, J., and H. Ashraf, "Advanced Silicon Etching Using High Density Plasmas," *Proc. SPIE, Micromachining and Microfabrication Process Technology Symp*., Austin, TX, October 23–24, 1995, Vol. 2639, pp. 224–233.

[15] Lärmer, F., and P. Schilp, "Method of Anisotropically Etching Silicon," German Patent DE 4 241 045, 1994.

[16] Ayón, A. A., et al., "Etching Characteristics and Profile Control in a Time Multiplexed Inductively Coupled Plasma Etcher," *Tech. Digest Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, June 8–11, 1998, pp. 41–44.

[17] Lasky, J. B., "Wafer Bonding for Silicon-On-Insulator Technologies," *Applied Physics Letters*, Vol. 48, No. 1, January 6, 1986, pp. 78–80.

[18] Petersen, K. E., et al., "Silicon Fusion Bonding for Pressure Sensors," *Tech. Digest Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, June 6–9, 1988, pp. 144–147.

[19] Schmidt, M.A, "Wafer-to-Wafer Bonding for Microstructure Formation," *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1575–1585.

[20] Tong, Q.–Y., and U. Gösele, *Semiconductor Wafer Bonding*, New York: Wiley, 1999, pp. 49–72.

[21] Miki, N., et al., "Multi-Stack Silicon-Direct Wafer Bonding for 3D MEMS Manufacturing," *Sensors and Actuators A*, Vol. 103, 2003, pp. 191–194.

[22] Strawbridge, I., and P. F. James, "Glass Formation from Gels," in *High Performance Glasses*, M. Cable and J. M. Parker (eds.), London: Blackie Publishing, 1992, pp. 20–49.

[23] Guckel, H., "High-Aspect Ratio Micromachining Via Deep X-Ray Lithography," *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1586–1593.

[24] Mulhern, G. T., D. S. Soane, and R. T. Howe, "Supercritical Carbon Dioxide Drying of Microstructures," *Proc. 7th Int. Conf. on Solid-State Sensors and Actuators (Transducers '93)*, Yokohama, Japan, June 7–10, 1993, pp. 296–299.

[25] Srinivasan, U., et al., "Lubrication of Polysilicon Micromechanisms with Self-Assembled Monolayers," *Technical Digest of Solid State Sensor and Actuator Workshop*, Hilton Head, SC, June 1998, pp. 156–161.

[26] MicroChem Corp., Data Sheet for NANO SU-8 2000, Newton, MA, February 2002.

[27] Mikroglas Teknik AG, "FOTURAN—A Photostructurable Glass," Mainz, Germany, 2003.

[28] Ehrfeld, W., and U. Ehrfeld, "Progress and Profit Through Micro Technologies: Commercial Applications of MEMS / MOEMS," *Proc. SPIE*, Vol. 4,557, October 2001, pp. 1–10.

[29] Kalpakjian, S., and S. R. Schmid, *Manufacturing Processes for Engineering Materials, 4th Edition*, Upper Saddle River, NJ: Pearson Education/Prentice Hall, 2003.

[30] DuPont Microcircuit Materials, "DuPont Green Tape™ Design and Layout Guideline," Research Triangle Park, NC, 2003.

[31] Kulke, R., et al., "LTCC—Multilayer Ceramic for Wireless and Sensor Applications," English translation of "LTCC-Mehrlagenkeramik für Funk- und Sensor-Anwendungen," *Produktion von Leiterplatten und Systemen (PLUS)*, Eugen G. Leuze Verlag, December 2001, pp. 2131–2136.

[32] Xia, Y., et al., "Non-Lithographic Methods for Fabrication of Elastomeric Stamps for Use in Microcontact Printing," *Langmuir*, Vol. 12, No. 16, 1996, pp. 4033–4038.

[33] B. Michel, et al., "Printing Meets Lithography: Soft Approaches to High-Resolution Patterning," *IBM Journal of Research and Development*, Vol. 45, No. 5, September 2001, pp. 697–714.

[34] Chou, S. Y., P. R. Krauss, and P. J. Renstrom, "Nanoimprint Lithography," *J. Vac. Sci. Technol. B*, Vol. 14, No. 6, November/December 1996, pp. 4129–4133.

[35] Bustillo, J. M., R. T. Howe, and R. S. Muller, "Surface Micromachining for Microelectromechanical Systems," *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1559–1561.

[36] Klaassen, E. H., et al., "Silicon Fusion Bonding and Deep Reactive Ion Etching; A New Technology for Microstructures," *Proc. 8th Int. Conf. on Solid-State Sensors and Actuators*, Stockholm, Sweden, June 25–29, 1995, pp. 556–559.

[37] Shaw, K. A., Z. L. Zhang, and N. C. MacDonald, "SCREAM-I: A Single Mask, Single-Crystal Silicon, Reactive Ion Etching Process for Microelectromechanical Structures," *Sensors and Actuators*, Vol. A40, No. 1, 1994, pp. 63–70.

Selected Bibliography

Chang, C. Y., and S. M. Sze (eds.), *ULSI Technology*, New York: McGraw-Hill, 1996.

Flamm, D. L., and G. K. Herb, "Plasma Etching Technology," in *Plasma Etching: An Introduction*, D. M. Manos and D. L. Flamm (eds.), San Diego, CA: Academic Press, 1989, pp. 1–89.

Jaeger, R. C., *Introduction to Microelectronic Fabrication*, Reading, MA: Addison-Wesley, 1988.

Kamins, T., *Polycrystalline Silicon for Integrated Circuits*, Boston, MA: Kluwer Academic Publishers, 1988.

Kovacs, G. T. A., *Micromachined Transducers Sourcebook*, New York: McGraw-Hill, 1998.

Madou, M., *Fundamentals of Microfabrication*, Boca Raton, FL: CRC Press, 1997.

Moreau, M., *Semiconductor Lithography Principles, Practices and Materials,* New York: Plenum Press, 1988.

Tong, Q.–Y., and U. Gösele, *Semiconductor Wafer Bonding*, New York: Wiley, 1999.

Wise, K. D. (ed.), "Special Issue on Integrated Sensors, Microactuators, and Microsystems (MEMS)," *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998.

CHAPTER 4

MEM Structures and Systems in Industrial and Automotive Applications

> "…for I believe that his device had tremendous advantages and unless there be other systems of equal merits which are unknown to me, I am of the opinion that he has the most remarkable system in existence."
>
> —*David Sarnoff on E. Howard Armstrong's radio receiver, 1914. Quoted in the* Empire of the Air*, by Tom Lewis.*

Armed with an understanding of the fabrication methods, it is time to examine various types of microelectromechanical (MEM) structures and systems. It is apparent that with a vast and diverse set of fabrication tools, creativity abounds. Indeed, the list of MEM structures and devices continues to grow daily as more applications prove to benefit from miniaturization. But just as necessity is mother of all inventions, it is economics that ultimately determine the commercial success of a particular design or technology. Demonstrations of micromachined devices are innumerable, but the successful products are few. MEMS technology is only a means to achieve a solution for a particular application. A quest for its perfection should not entail an oversight of the end objective: the application itself.

The next four chapters review a select set of MEMS-based commercial products with applications in multiple diverse markets. This chapter is specific to those products with utility in industrial and automotive applications. It also includes a short introduction on the general methodology of the design process and a listing of commonly used sensing and actuation techniques. The following three chapters address products for optical, life sciences, and RF electronic applications.

Three general categories form the total extent of MEMS: *sensors*, *actuators*, and *passive structures*. Sensors are transducers that convert mechanical, thermal, or other forms of energy into electrical energy; actuators do the exact opposite. Passive structures include devices where no transducing occurs, including both mechanical and optical components. A complete listing of all MEMS demonstrations is not sought in this book; rather, the theme is to illustrate the state of the technology by providing sufficient examples of structures and systems that have proven their commercial viability or show promise to do so in the near future.

General Design Methodology

Starting with a list of specifications for the MEM device or system, the design process begins with the identification of the general operating principles and overall

structural elements, then proceeds onto analysis and simulation, and finally onto outlining of the individual steps in the fabrication process. This is often an iterative process involving continuous adjustments to the shape, structure, and fabrication steps. The layout of the lithographic masks is the final step before fabrication and is completed using specialized CAD tools to define the two-dimensional patterns.

Early design considerations include the identification of the general sensing or actuation mechanisms based on performance requirements. For instance, the output force requirement of a mechanical microactuator may favor thermal or piezoelectric methods and preclude electrostatic actuation. Similarly, the choice of piezoresistive sensing is significantly different from capacitive or piezoelectric sensing. The interdisciplinary nature of the field brings together considerations from a broad range of specialties, including mechanics, optics, fluid dynamics, materials science, electronics, chemistry, and even biological sciences. On occasion, determining a particular approach may rely on economic considerations or ease of manufacture rather than performance. For example, the vast majority of pressure sensors use cost-effective piezoresistive sense elements instead of the better performing, but more expensive, resonant-type sense structures.

The design process is not an exact analytical science but rather involves developing engineering models, many for the purpose of obtaining basic physical insights. Computer-based simulation tools using finite-element modeling are convenient for analyzing complex systems. A number of available programs, such as ANSYS® (ANSYS, Inc., of Canonsburg, Pennsylvania) and CoventorWare™ (Coventor, Inc., of Cary, North Carolina), can simulate mechanical, thermal, and electrostatic structures (see Figure 4.1). Substantial efforts are currently under way to develop sophisticated programs that can handle coupled multimode problems, (e.g., simultaneously combining fluid dynamics with thermal and mechanical analysis). As powerful as these tools are perceived to be, their universal predictive utility is questionable. However, they can provide valuable insight into and visualization of the device's operation.

In planning a fabrication process, the choice is to use a standard foundry service with a completely predefined process flow, to use a service that allows the selection

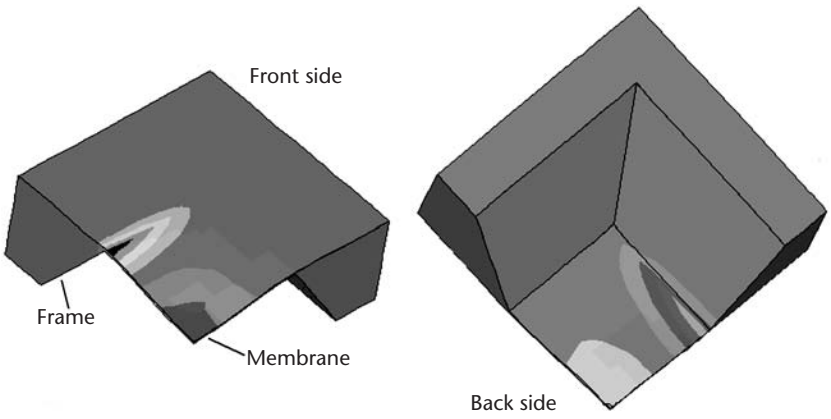


Figure 4.1 A finite element simulation using ANSYS modeling program of a quarter of a bulk micromachined silicon pressure sensor showing contours of mechanical stress in response to an applied pressure load.

of previously developed individual process steps, or to design a custom process specific to the device or system. If the production unit volume is not sufficiently large, it may be challenging to identify reputable manufacturing facilities willing to develop and implement custom processes.

Techniques for Sensing and Actuation

Common Sensing Methods

Sensing is by no means a modern invention. There are numerous historical accounts describing the measurement of physical parameters—most notably, distance, weight, time, and temperature. Early Chinese attempts at making compasses date back to the twelfth century with the use of lodestone, a naturally occurring magnetic ore. Modern sensing methods derive their utility from the wealth of scientific knowledge accumulated over the past two centuries. We owe our intimate familiarity with electrostatics and capacitance to the work of Charles Augustin de Coulomb of France and John Priestly of England in the late eighteenth century and observe that Lord Kelvin's discovery of piezoresistivity in 1856 is recent in historical terms. What distinguishes these modern techniques is the ability to sense with greater accuracy and stability; what makes them suitable for MEMS is their scalable functionality.

The objective of modern sensing is the transducing of a specific physical parameter, to the exclusion of other interfering parameters, into electrical energy. Occasionally, an intermediate conversion step takes place. For example, pressure or acceleration are converted into mechanical stress, which is then converted to electricity. Infrared radiation in image sensors is often converted into heat and then sensed as an electrical voltage or a change in electrical resistance. Perhaps the most common of all modern sensing techniques is temperature measurement using the dependence of various material properties on temperature. This effect is pronounced in the electrical resistance of metals. The rate at which the resistance rises with temperature—TCR—of most metals ranges between 10 and 100 parts per million per degree centigrade.

Piezoresistivity and piezoelectricity are two sensing techniques described in greater detail in Chapter 2 (see Table 4.1). Impurity-doped silicon exhibits a piezoresistive behavior that lies at the core of many pressure and acceleration sensor

Table 4.1 The Relative Merits of Piezoresistive, Capacitive, and Electromagnetic Sensing Methods

Piezoresistive	*Capacitive*	*Electromagnetic*
Simple fabrication	Simple mechanical structure	Structural complexity varies
Low cost	Low cost	Complex packaging
Voltage or current drive	Voltage drive	Current drive
Simple measurement circuits	Requires electronic circuits	Simple control circuits
Low-output impedance	Susceptible to EMI	Susceptible to EMI
High-temperature dependence	Low-temperature dependence	Low temperature dependence
Small sensitivity	Large dynamic range	Sensitivity $\propto$ magnetic field
Insensitive to parasitic resistance	Sensitive to parasitic capacitance	Insensitive to parasitic inductance
Open loop	Open or closed loop	Open or closed loop
Medium power consumption	Low power consumption	Medium power consumption

designs. Measuring the change in resistance and amplifying the corresponding output signal tend to be rather simple, requiring a basic knowledge of analog circuit design. A drawback of silicon piezoresistivity is its strong dependence on temperature which must be compensated for with external electronics.

By contrast, capacitive sensing relies on an external physical parameter changing either the spacing or the relative dielectric constant between the two plates of a capacitor. For instance, an applied acceleration pushes one plate closer to the other. Or in the example of relative humidity sensors, the dielectric is an organic material whose permittivity is function of moisture content [1]. The advantages of capacitive sensing are very low power consumption and relative stability with temperature. Additionally, the approach offers the possibility of electrostatic actuation to perform closed-loop feedback. The following section on actuation methods explains this point further. Naturally, capacitive sensing requires external electronics to convert minute changes in capacitance into an output voltage. Unlike measuring resistance, these circuits can be substantially intricate if the change in capacitance is too small. This is frequently the case in MEMS where capacitance values are on the order of 1 pF (10^{-12} F) and changes in capacitance can be as small as a few fF (10^{-15} F).

Yet another sensing approach utilizes electromagnetic signals to detect and measure a physical parameter. Magnetoresistive sensors on the read heads of high-density computer disk drives measure the change in conductivity of a material slab in response to the magnetic field of the storage bit. In Hall-effect devices, a magnetic field induces a voltage in a direction orthogonal to current flow [2]. Hall-effect sensors are extremely inexpensive to manufacture. They are used in high-reliability computer keyboards and make excellent candidates to measure wheel velocity in vehicles. Another form of electromagnetic transducing uses Faraday's law to detect the motion of a current-carrying conductor through a magnetic field. Two yaw-rate sensors described later in this chapter make use of this phenomenon. The control electronics for magnetic sensors can be readily implemented using modern CMOS technology, but generating magnetic fields often necessitates the presence of a permanent magnet or a solenoid.

Common Actuation Methods

A complete shift in paradigm becomes necessary to think of actuation on a miniature scale—a four-stroke engine is not scalable. The next five schemes illustrate the diversity and the myriad of actuation options available in MEMS. They are *electrostatic*, *piezoelectric*, *thermal*, *magnetic*, and *phase recovery using shape-memory alloys*. The choice of actuation depends on the nature of the application, ease of integration with the fabrication process, the specifics of the system around it, and economic justification (see Table 4.2). Examples of each actuation method will arise throughout this chapter and the next.

Electrostatic actuation

Electrostatic actuation relies on the attractive force between two conductive plates or elements carrying opposite charges. A moment of thought quickly reveals that the charges on two objects with an *externally applied* potential between them can only be of opposite polarities. Therefore, an applied voltage, regardless of its polarity,

Table 4.2 Comparison of Various Actuation Methods on the Basis of Maximum Energy Density. Actual Energy Output May Be Substantially Lower Depending on the Overall Efficiency of the System

Actuation	*Max. Energy Density*	*Physical and Material Parameters*	*Estimated Conditions*	*Approximate Order (J/cm³)*
Electrostatic	$\frac{1}{2}\,\varepsilon_0 E^2$	E = electric field ε_0 = dielectric permittivity	5 V/μm	~ 0.1
Thermal	$\frac{1}{2}\,Y(\alpha\,\Delta T)^2$	α = coefficient of expansion ΔT = temperature rise Y = Young's modulus	3×10^{-6} /°C 100°C 100 GPa	~ 5
Magnetic	$\frac{1}{2}\,B^2/\mu_0$	B = magnetic field μ_0 = magnetic permeability	0.1 T	~ 4
Piezoelectric	$\frac{1}{2}\,Y(d_{33}\,E)^2$	E = electric field Y = Young's modulus d_{33} = piezoelectric constant	30 V/μm 100 GPa 2×10^{-12} C/N	~ 0.2
Shape-memory alloy	—	Critical temperature		~ 10 (from reports in literature)

always results in an attractive electrostatic force. If C is the capacitance between two parallel plates [see Figure 4.2(a)], x is the spacing between them, and V is an externally applied voltage, the electrostatic force is then $\frac{1}{2}CV^2/x$ (the square term ensures that the force is always positive and attractive). For two parallel plates with a spacing of one micrometer, an applied voltage of 5V, and a reasonable area of 1,000 μm^2, the electrostatic force is merely 0.11 μN. Electrostatic comb actuators [3] are a variant that includes two comb sets of interdigitated "teeth" that are offset relative to each other [see Figure 4.2(b)]. An applied voltage brings the two combs together such that the teeth become alternating. Designers have favored comb actuators over parallel-plate actuators for two primary reasons: they allow a larger displacement (tens of micrometers are feasible) and the force is relatively independent of displacement. Forces are, however, of the same order as forces for a parallel plate with the same quadratic dependence on voltage.

A natural extension of electrostatic actuation is closed-loop feedback in systems employing capacitive sensing. When sense circuits detect the two surfaces of a capacitor separating under the effect of an external force (e.g., acceleration), an electrostatic feedback voltage is immediately applied by the control electronics to counteract the disturbance and maintain a fixed capacitance. The magnitude of the

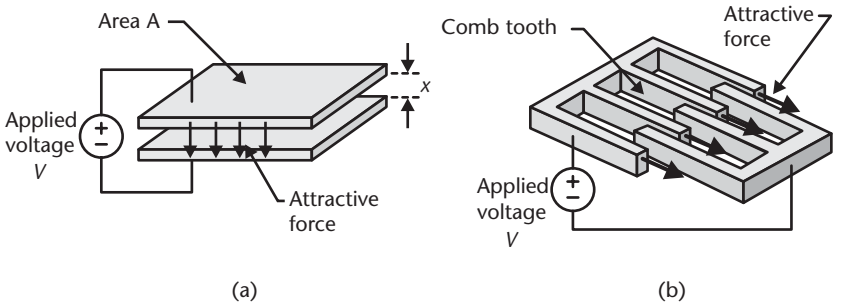


Figure 4.2 (a) An illustration of a parallel-plate electrostatic actuator with an applied voltage V and a spacing x. The attractive force is normal to the plate surfaces. (b) An illustration of an electrostatic comb actuator. The attractive force is in the direction of the interdigitated teeth.

feedback voltage then becomes a measure of the disturbing force. This feature is integral to the closed-loop operation of many accelerometers and yaw-rate sensors.

Piezoelectric Actuation

Piezoelectric actuation can provide significantly large forces, especially if thick piezoelectric films are used. Commercially available piezoceramic cylinders can provide up to a few newtons of force with applied potentials on the order of a few hundred volts. However, thin-film ($<5\ \mu$m) piezoelectric actuators can only provide a few millinewtons. Both piezoelectric and electrostatic methods offer the advantage of low power consumption as the electric current is very small.

Thermal Actuation

Thermal actuation consumes more power than electrostatic or piezoelectric actuation but can provide, despite its gross inefficiencies, actuation forces on the order of hundreds of millinewtons or higher. At least three distinct approaches have emerged within the MEMS community. The first capitalizes on the difference in the coefficients of thermal expansion between two joined layers of dissimilar materials to cause bending with temperature—the classic case of a bimetallic thermostat studied by S. Timoshenko in 1925 [4]. One layer expands more than the other as temperature increases. This results in stresses at the interface and consequently bending of the stack. The amount of bending depends on the difference in coefficients of thermal expansion and absolute temperature. Unfortunately, the latter dependence severely limits the operating temperature range—otherwise, the device may actuate prematurely on a hot day.

In another approach known as thermopneumatic actuation, a liquid is heated inside a sealed cavity. Pressure from expansion or evaporation exerts a force on the cavity walls, which can bend if made sufficiently compliant. This method also depends on the absolute temperature of the actuator. Valves employing this method will be described later in this chapter.

Yet a third distinct method utilizes a suspended beam of a same homogeneous material with one end anchored to a supporting frame of the same material [5]. Heating the beam to a temperature above that of the frame causes a differential elongation of the beam's free end with respect to the frame. Holding this free end stationary gives rise to a force proportional to the beam length and temperature differential. Such an actuator delivers a maximal force with zero displacement, and conversely, no force when the displacement is maximal. Designs operating between these two extremes can provide both force and displacement. A system of mechanical linkages can optimize the output of the actuator by trading off force for displacement, or vice versa. Actuation in this case is independent of fluctuations in ambient temperature because it relies on the difference in temperature between the beam and the supporting frame. A plate microvalve utilizing this actuation scheme is described later.

Magnetic Actuation

Lorentz forces form the dominant mechanism in magnetic actuation on a miniaturized scale [6]. This is largely due to the difficulty in depositing permanently

magnetized thin films. Electrical current in a conductive element that is located within a magnetic field gives rise to an electromagnetic force—the Lorentz force—in a direction perpendicular to the current and magnetic field. This force is proportional to the current, magnetic flux density, and length of the element. A conductor 1 mm in length carrying 10 mA in a 1-T magnetic field is subject to a force of 10 μN. Lorentz forces are useful for closed-loop feedback in systems employing electromagnetic sensing. Two yaw-rate sensors and a beam steering micromirror described later make use of this method.

Actuation Using Shape-Memory Alloys

Finally, of all five schemes, shape-memory alloys undoubtedly offer the highest energy density available for actuation. The effect, introduced in Chapter 2, can provide very large forces when the temperature of the material rises above the critical temperature, typically around 100°C. The challenge with shape-memory alloys lies in the difficulty of integrating their fabrication with conventional silicon manufacturing processes.

Passive Micromachined Mechanical Structures

Fluid Nozzles

Nozzles are among the simplest microstructures to fabricate using anisotropic etching of silicon, electroforming, or laser drilling of a metal sheet. A series of U.S. patents issued in the 1970s to IBM Corp. [7] describes the fabrication of silicon nozzles and their application for inkjet printing. The Ford Motor Company experimented in the 1980s with micromachined nozzles for engine fuel injection. With the expiration of most key patents on nozzle formation, micromachined nozzles are becoming common features in the design of atomizers, medical inhalers, and fluid spray systems. Nozzles need not necessarily be of silicon. MicroParts GmbH of Dortmund, Germany, manufactures a drug-inhaling device for asthma patients that incorporates a precise plastic nozzle fabricated using the LIGA electroplating and molding process described in the previous chapter.

A simple square silicon nozzle can be readily fabricated by depositing silicon nitride on both sides of a (100) wafer and patterning a square in the silicon nitride layer on the back side. Anisotropic etching in potassium hydroxide (KOH) or tetramethyl ammonium hydroxide (TMAH) forms a port through the wafer with walls defined by the {111} planes of silicon. The dimensions of the backside opening in the silicon nitride must be larger than 71% of the wafer thickness in order to etch through the wafer (see Figure 4.3).

Forming nozzles of circular or arbitrary shape in silicon involves additional fabrication steps. The most common approach is to grow on a (100) wafer a *p*-type epitaxial layer of silicon with a high boron concentration ($>1 \times 10^{19}$ cm^{-3}). The shape of the nozzle is patterned and etched into the *p*-type silicon layer using standard lithography and plasma etching (or RIE). A protective layer of silicon nitride is deposited on both sides of the wafer and patterned in the shape of a square on the back side. Double-sided lithography provides accurate alignment between the nozzle opening and the square on the back side. The fabrication is complete with the

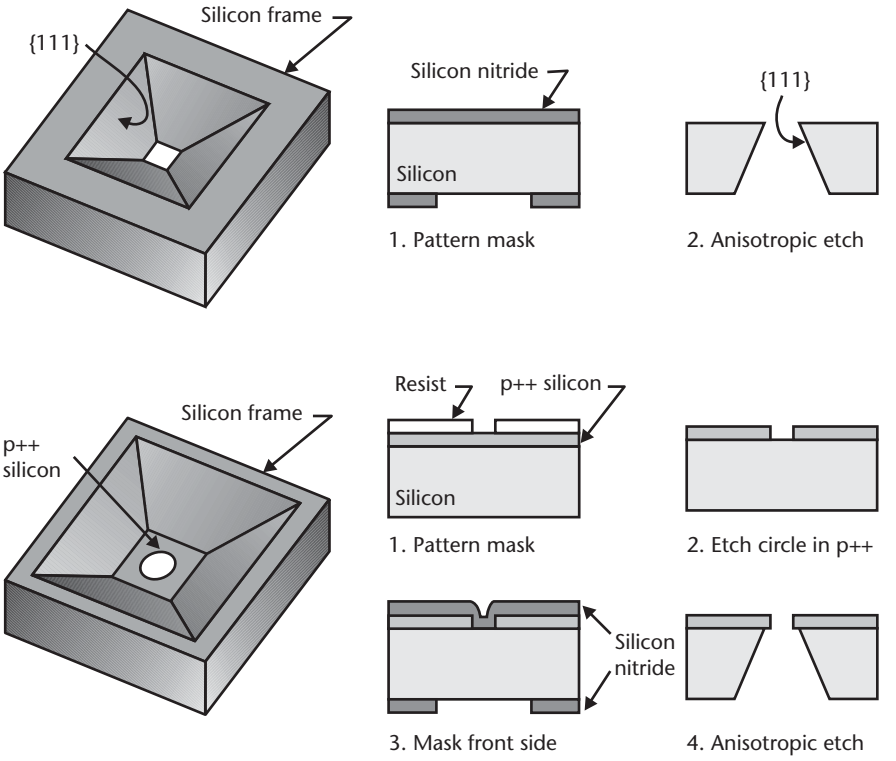


Figure 4.3 Schematic illustrations of square and circular nozzles on the wafer surface with their corresponding fabrication steps.

anisotropic etching of the silicon from the back side using KOH or TMAH. The *p*-type layer acts as an etch stop, thus preserving the shape of the nozzle.

These nozzles are oriented perpendicular to the surface of the wafer and are referred to as *top shooters* or *roof shooters* in the inkjet field. Nozzles oriented parallel to the wafer surface are termed *side shooters*. One such implementation developed by Xerox Corp. of Webster, New York, uses orientation-dependent etching to form grooves in a silicon wafer [8]. Another wafer is coated with a polyimide spacer layer and bonded to the grooved wafer. Finally, the wafer is diced to reveal triangle-shaped ports [Figure 4.4(a)].

Choosing a fluid flow path in the plane of the silicon wafer and using RIE provides further flexibility in shaping the nozzle and the orifices. In an implementation of a CO_2 cleaning apparatus [9], a silicon micromachined nozzle was specially designed to allow subsonic fluid flow at the inlet and supersonic flow at the outlet [Figure 4.4(b)]. DRIE is a suitable process for defining in the silicon a deep channel (50 to 500 μm) following the desired contour of the nozzle. The dimensional control is limited in the plane of the wafer by the lithography to better than one micrometer, whereas in the vertical depth direction, it is limited by the etch process to approximately 10% of the total depth. A top cover is later bonded using anodic bonding of glass or silicon fusion bonding.

Nozzles can alternatively be fabricated using electroforming. The process starts with the production of a mold or *mandrel*, which may be flat or have topography

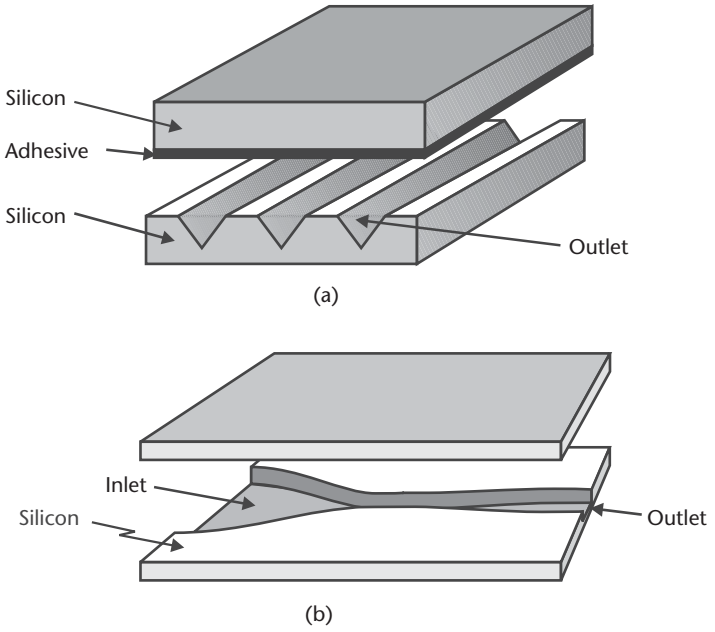


Figure 4.4 Illustration of side-shooter nozzles: (a) nozzles formed by orientation-dependent etching of grooves, wafer bonding, and dicing [8], and (b) nozzle formed by DRIE and wafer bonding. (*After:* [9].)

such as bumps and trenches (see Figure 4.5). The mandrel material must be electrically conducting to enable electroplating and have sufficient adhesion to the metal being plated but allow the metal to be peeled off after plating [10]. For example, the materials system used by Hewlett-Packard for early-generation inkjet orifice plates is electroplated nickel on a stainless steel mandrel: stainless steel has a thin epitaxial oxide layer on it that allows electrical conduction but does not form a strong bond to the plated nickel. Photoresist or some other insulator is patterned on the mandrel

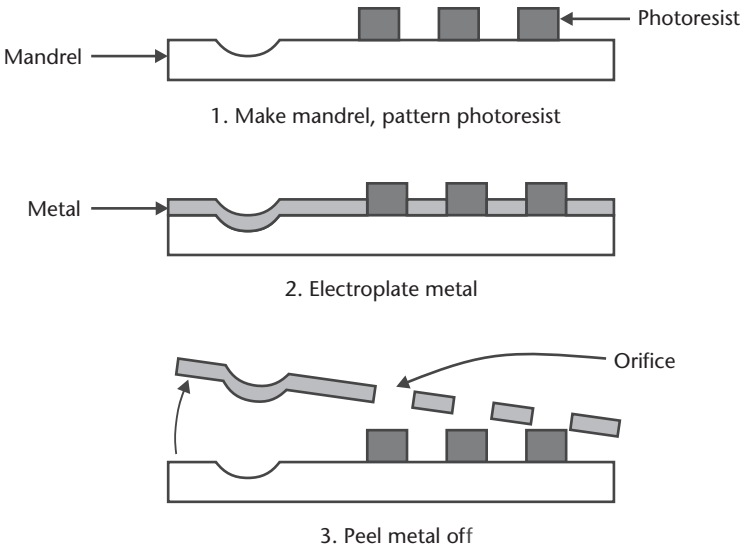


Figure 4.5 Illustration of an electroformed nozzle process.

surface where through-holes are desired. Metal is then electroplated everywhere in the mandrel that is not protected by the photoresist. Finally, the plated metal-foil structure is peeled off of the mandrel and the resist is stripped. A later section in this chapter describes the inkjet head in greater detail.

Hinge Mechanisms

Hinges are very useful passive elements in our daily lives. At the microscopic scale, they extend the utility of the inherently two-dimensional surface micromachining technology into the third dimension. The hinge fabrication occurs simultaneously with the rest of the planar structures on the wafer (see Figure 4.6). Folding the hinge out of the plane gives structures access to the space above the silicon die. One potential future commercial application that may benefit from these fold-up mechanisms is the assembly of microlenses, mirrors, and other components on optical microbenches [11, 12] (see Figure 4.7).

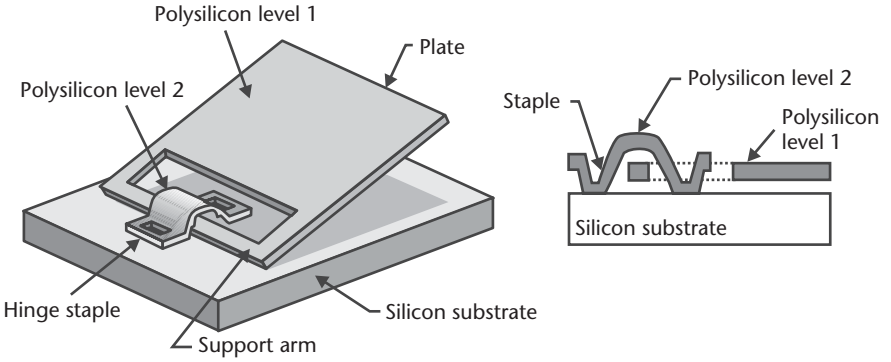


Figure 4.6 Illustration of the fold-up surface micromachined hinge. The structure is fabricated using polysilicon surface micromachining. (*After:* [13].)

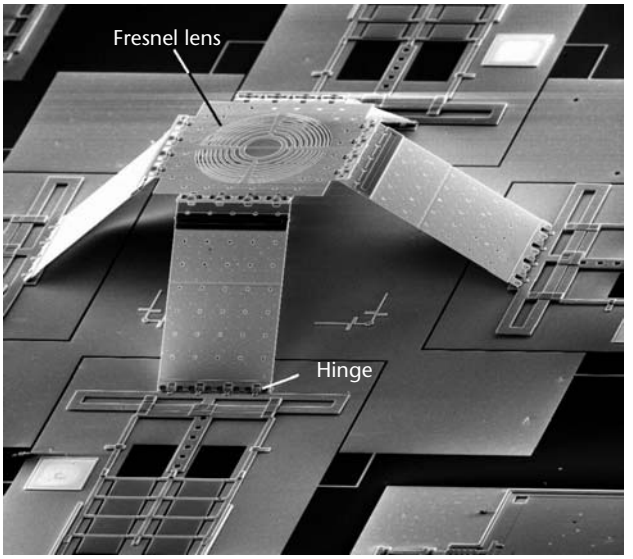


Figure 4.7 Photograph of a Fresnel microlens on an adjustable platform made of five hinged polysilicon plates. (*Courtesy of*: M. Wu, University of California, Los Angeles.)

The hinge structure is simple, consisting of a plate and a support arm made of a first polysilicon layer. A staple made of a second polysilicon layer captures the plate support arm. The staple is anchored directly to the substrate. The fabrication utilizes the polysilicon surface micromachined process introduced in Chapter 3. The polysilicon layers are typically 2 μm thick. The sacrificial phosphosilicate glass (PSG) layer is 0.5 to 2.5 μm thick. Etching in hydrofluoric acid removes the PSG layer and releases the mechanical plate from the substrate. Recent designs incorporate mechanical levers that snap into grooves defined in the plate and permanently lock the hinge in a vertical position.

In early demonstrations, the assembly process involved manually lifting each plate into position using sharp probes. The process has recently evolved to rely on self assembly by designing the hinges so that they lock in place when the movable parts are at a particular angle relative to the substrate. Random agitation while rinsing in water swings the structures away from the substrate; when they reach a preset design location, they latch and lock in position. Both manual and self-assembly tasks remain tedious and must be automated in the future before hinge assembly gains acceptance in a mainstream manufacturing environment.

Sensors and Analysis Systems

Pressure Sensors

The first high-volume production of a pressure sensor began in 1974 at National Semiconductor Corp. of Santa Clara, California. Pressure sensing has since grown to a large market with an estimated 60 million silicon micromachined pressure sensors manufactured in 2001. Nearly all units use bulk micromachining technology. Manifold-absolute-pressure (MAP) [14] and disposable blood pressure [15] sensing are the two single largest applications. The vast majority use piezoresistive sense elements to detect stress in a thin silicon diaphragm in response to a pressure load. A few designs use capacitive methods to sense the displacement of a thin diaphragm.

The basic structure of a piezoresistive pressure sensor consists of four sense elements in a Wheatstone bridge configuration that measure stress within a thin crystalline silicon membrane (see Figure 4.8). The stress is a direct consequence of the membrane deflecting in response to an applied pressure differential across the front and back sides of the sensor. The stress is, to a first order approximation, linearly proportional to the applied pressure differential. The membrane deflection is typically less than one micrometer. The output at full-scale applied pressure is a few millivolts per volt of bridge excitation (the supply voltage to the bridge). The output normalized to input applied pressure is known as sensitivity [(mV/V)/Pa] and is directly related to the piezoresistive coefficients, $\pi_{//}$ and $\pi_{\perp}$ (see Chapter 2). The thickness and geometrical dimensions of the membrane affect the sensitivity and, consequently, the pressure range of the sensor. Devices rated for very low pressures (less than 10 kPa) usually incorporate complex membrane structures, such as central bosses, to concentrate the stresses near the piezoresistive sensors and improve both sensitivity and linearity.

A common design layout on {100} substrates positions the four diffused *p*-type piezoresistors at the points of highest stress, which occur at the center edges of the

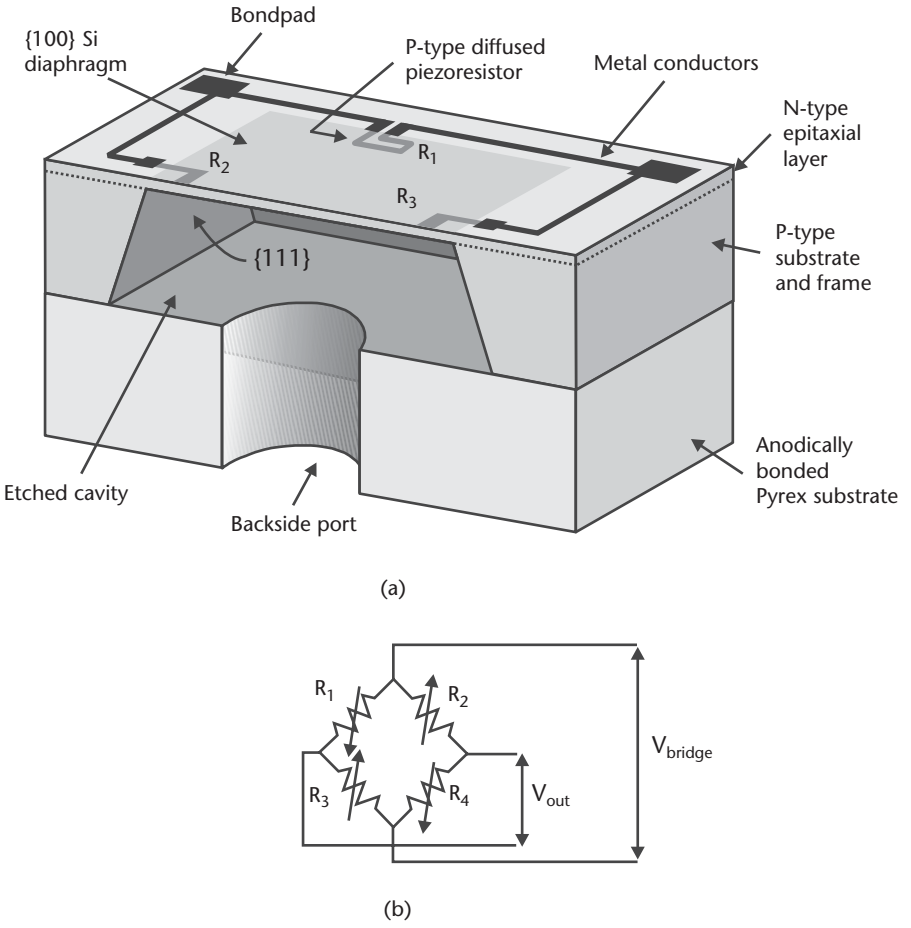


Figure 4.8 (a) Schematic illustration of a pressure sensor with diffused piezoresistive sense elements; and (b) the four sense elements form a Wheatstone bridge configuration.

diaphragm. Two resistors have their primary axes parallel to the membrane edge, resulting in a decrease in resistance with membrane bending. The other two resistors have their axes perpendicular to the edge, which causes the resistance to increase with the pressure load. Other layouts are also possible including designs to measure shear stress, but the main objective remains to position the resistors in the areas of highest stress concentration in order to maximize the response to applied pressure. It is necessary that the four piezoresistors have identical resistances in the absence of applied pressure. Any mismatch in resistance, even one caused by temperature, causes an imbalance in the Wheatstone bridge. The resulting output reading is known as zero offset and is undesirable.

Deep diffusions degrade the sensitivity of the piezoresistors by averaging the stress over the depth of the sense element. Shallow diffusions are prone to surface charge effects that can cause long-term drift in the output signal. Remedies to these conflicting requirements are frequently proprietary to the manufacturers. U.S. patent 4,125,820 (November 14, 1978) assigned to Honeywell, Inc., of Minneapolis, Minnesota, illustrates one solution in which the piezoresistive diffusions are buried below the surface of the membrane.

The fabrication process of a typical pressure sensor relies mostly on steps standard to the integrated circuit industry, with the exception of the precise forming of the thin membrane using electrochemical etching (ECE). An *n*-type epitaxial layer of silicon is grown on a *p*-type, {100} wafer. A thin, preferably stress-free, insulating layer is deposited or grown on the front side of the wafer, and a protective silicon nitride film is deposited on the back side. The piezoresistive sense elements are formed by locally doping the silicon *p*-type using the masked implantation of boron, followed by a high-temperature diffusion cycle. Etching of the insulator on the front side provides contact openings to the underlying piezoresistors. A metal layer, typically aluminum, is then sputter deposited and patterned in the shape of electrical conductors and bond pads. A square opening is patterned and etched in the silicon nitride layer on the back side. Double-sided lithography ensures that the backside square is precisely aligned to the sense elements on the front side. At this point, electrical contacts are made to the *p*-type substrate and *n*-type epitaxial layer, and the silicon is electrochemically etched from the back side in a solution of potassium hydroxide. Naturally, the front side must be protected during the etch. One practical protection method includes coating with wax such as paraffin and clamping in a fixture. The etch stops as soon as the *p*-type silicon is completely removed, and the *n*-type layer is exposed. The process forms a membrane with precise thickness defined by the epitaxial layer. Anodic bonding in vacuum of a Pyrex glass wafer on the back side produces an absolute pressure sensor that measures the pressure on the front side in reference to the cavity pressure (often, vacuum). For differential- or gauge-type pressure sensors, previously drilled holes in the glass wafer provide vent ports (see Figure 4.9).

The advent of silicon-fusion bonding in the 1980s proved very useful to the design of bulk micromachined pressure sensors. The outward sloping of {111} planes delineating the sensor's frame results in an unnecessary increase in die size. Silicon-fusion bonding allows the forming of the membrane after the etching of a reference cavity with inward sloping {111} walls. Consequently, extremely small pressure sensors are feasible. For example, GE NovaSensor of Fremont, California, manufactures a sensor that is 400 μm wide, 800 μm long, and 150 μm thick, and it fits inside the tip of a catheter (see Figure 4.10).

The fabrication of a silicon-fusion-bonded sensor begins with the etching of a cavity in a bottom handle wafer. Silicon-fusion bonding of a *p*-type top wafer with an *n*-type epixatial layer encapsulates and seals the cavity. Electrochemical etching or standard polishing thins down the top bonded wafer to form a membrane of appropriate thickness. The remaining process steps define the piezoresistive sense elements, as well as the metal interconnects, and are similar to those used in the fabrication of the standard bulk micromachined pressure sensors described earlier.

Calibration and correction of error sources are necessary for the manufacture of precision pressure sensors. A specification on accuracy of better than 1% over a temperature range of –40° to +125°C is typical of many automotive, medical, and industrial applications. First-order errors include zero offset (the output at no applied pressure), uncalibrated sensitivity or span (conversion factor between input pressure and output signal), and temperature dependence of the output signal. Second-order effects include nonlinearities in the output response, as well as temperature coefficients of some first-order error terms. Compensation and correction

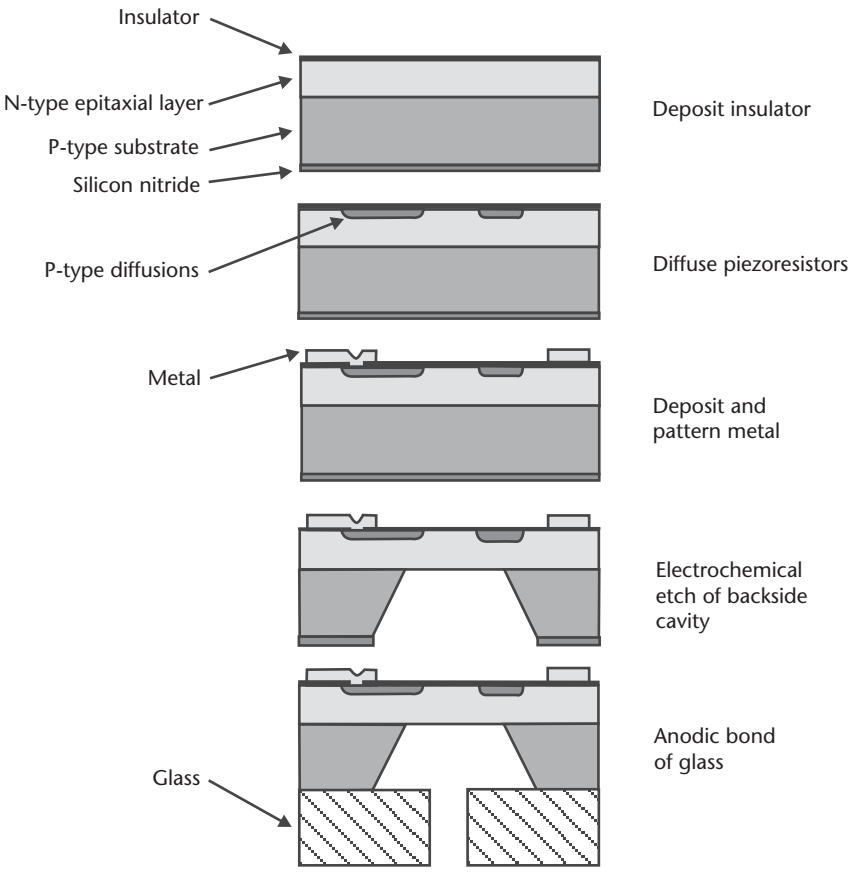


Figure 4.9 Fabrication steps for a piezoresistive, gauge, or differential bulk micromachined pressure sensor.

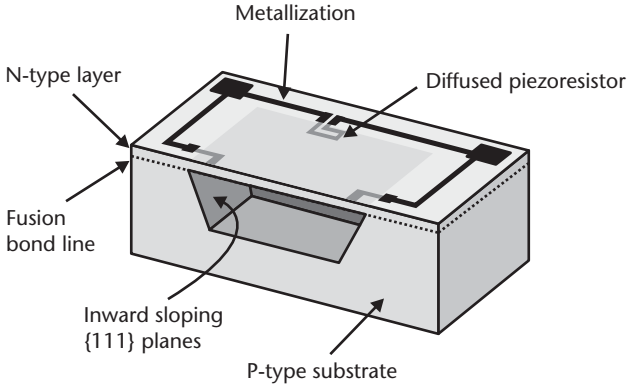


Figure 4.10 A miniature silicon-fusion-bonded absolute pressure sensor. (*Courtesy of:* GE NovaSensor of Fremont, California.)

techniques place certain restrictions on the device and process design. For example, one scheme uses laser trimming of resistors with near-zero TCR to correct first-order errors, but this scheme requires that the average doping concentration of the

piezoresistors is above 5×10^{19} cm^{-3}, or below 3×10^{17} cm^{-3}. Compensation over intermediate ranges of dopant concentration requires sophisticated electronic circuits that continuously monitor the temperature of the Wheatstone bridge.

There has been recently a shift within the industry to provide the calibration and compensation functions with specially designed application-specific integrated circuits (ASICs). The active circuits amplify the voltage output of the piezoresistive bridge to standard CMOS voltage levels (0–5V). They also correct for temperature errors and nonlinearities. Error coefficients particular to individual sensors are permanently stored in on-board electrically programmable memory (e.g., EEPROM). Most sensor manufacturers have developed their own proprietary circuit designs, and some have even integrated the circuitry onto the pressure-sensor chip. A few general-purpose signal conditioning integrated circuits are commercially available; one example is the MAX1457 from Maxim Integrated Products of Sunnyvale, California.

High-Temperature Pressure Sensors

The temperature rating of most commercially available silicon micromachined pressure sensors is –40° to +125°C, covering the automotive and military specifications. The increased leakage current above 125°C across the *p-n* junction between the diffused piezoresistive element and the substrate significantly degrades performance. Silicon-on-insulator (SOI) technology becomes very useful at elevated temperatures because the thin silicon sense elements exist over a layer of silicon dioxide, thus eliminating all *p-n* diode junctions. Adjacent silicon sense elements are isolated from each other by shallow moat-like trenches. The dielectric isolation below the sense elements completely eliminates the leakage current through the substrate, as long as the applied voltages are below the breakdown voltage of the insulating oxide layer.

A high-temperature pressure sensor from GE NovaSensor utilizes SOI technology to form thin *p*-type crystalline silicon piezoresistors over a thin layer of silicon dioxide. Gold metallization and bond pads provide electrical contacts to the sense elements (see Figure 4.11).

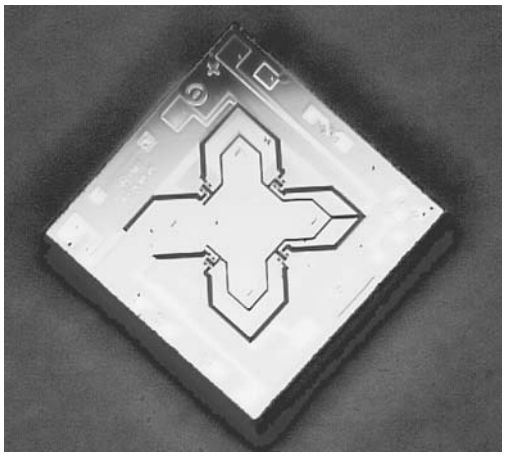


Figure 4.11 Photograph of an SOI-based pressure sensor rated for extended temperature operation up to 300°C. (*Courtesy of:* GE NovaSensor of Fremont, California.)

Silicon-fusion bonding plays an important role in the making of the SOI substrates (see Figure 4.12). A heavily doped, thin *p*-type layer is formed on the surface of one wafer, and an oxide layer is thermally grown on an another wafer. Silicon-fusion bonding brings the two substrates together such that the *p*-type layer is in direct contact with the oxide layer. Etching in ethylenediamine pyrocathechol (EDP) thins down the stack and stops on the heavily doped *p*-type silicon. A front-side lithography step followed by a silicon etch patterns the piezoresistive sense elements. Gold metallization is sputtered or evaporated and then lithographically patterned to form electrical interconnects and bond pads. The final step forms a thin membrane by etching a cavity from the back side using potassium hydroxide or a similar etch solution. Double-sided lithography is critical to align the cavity outline on the back side with the piezoresistors on the front side. The front side need not be protected during the etch of the cavity if EDP is used instead of potassium hydroxide; EDP is highly selective to heavily doped *p*-type silicon, silicon dioxide, and gold, but it is toxic and must be handled with extreme caution.

Mass Flow Sensors

The flow of gas over the surface of a heated element produces convective heat loss at a rate proportional to mass flow. Flow sensors operating on this principle belong to a general category of devices known as *hot-wire anemometers,* which measure the temperature of the hot element and infer the flow rate. A number of demonstrations exist in the open literature; most share a basic structure consisting of a thin-film heating element and a temperature-measuring device on a thin ($<1\ \mu$m) insulating dielectric membrane suspended over an etched cavity, at least 50 μm in depth. This architecture provides excellent thermal isolation between the heater and the supporting mechanical frame, which ensures that heat loss is nearly all due to mass flow over the heating element. A thermal isolation exceeding 15°C per milliwatt of heater power is typical. Moreover, the small heat capacity due to the tiny heated volume

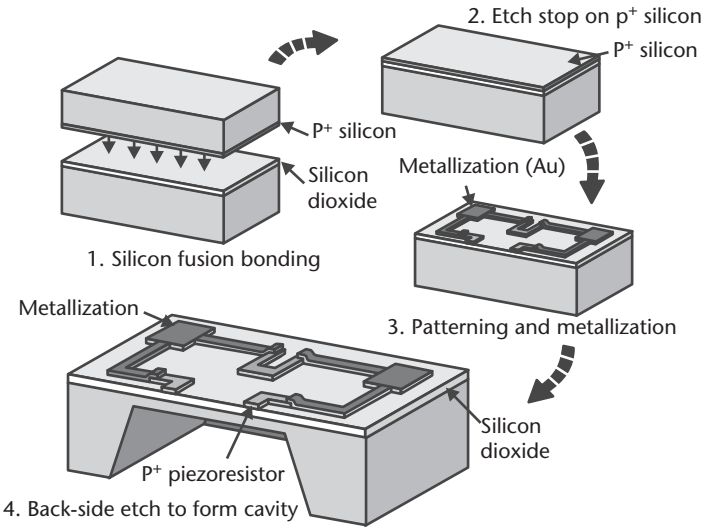


Figure 4.12 Fabrication process of an SOI high-temperature pressure sensor. (*Courtesy of:* GE NovaSensor of Fremont, California.)

provides a fast thermal time constant and consequently a rapid response time. One approach to inferring the temperature of the heating element is to drive it with a constant current and measure its resistance, then calculate the temperature using the TCR. Alternatively, direct temperature measurement using a *p-n* diode or a thermocouple is equally adequate.

Honeywell, Inc., of Minneapolis, Minnesota, manufactures the AWM series of bidirectional mass airflow sensors using two adjacent thin membranes, presumably made of silicon nitride, each containing a heating element and a temperature-sensitive resistor [16]. The two membranes are small in size, each measuring less than $500 \times 500\ \mu m^2$. Gas flow across the membranes cools the upstream heater and heats the downstream element. The two heaters are part of a first Wheatstone bridge, and the temperature-sensing resistors form two legs in a second Wheatstone bridge, whose differential output is directly proportional to the rate of flow (see Figure 4.13). The direction of flow is reflected in the polarity of the differential bridge output—a characteristic of the dual sense element configuration. In essence, this polarity determines which of the two heaters is upstream or downstream. Laser-trimmed thick- or thin-film resistors provide calibration as well as nulling of any offsets due to resistance mismatch in the Wheatstone bridges. The Honeywell AWM series of devices is capable of measuring gas flow rates in the range of 0 to 1,000 sccm. The upper limit is due to pronounced nonlinear effects in the heat-transfer mechanism. The full-scale output is approximately 75 mV, and the response time is less than 3 ms. The device consumes less than 30 mW.

While the processing details of the Honeywell series of airflow sensors are not publicly disclosed, one can readily design a process for fabricating a demonstration-type device. An example process would begin with the deposition of a thin layer of silicon nitride, approximately 0.5 μm in thickness, over a {100} silicon wafer. Silicon nitride is usually an excellent choice for making thin membranes

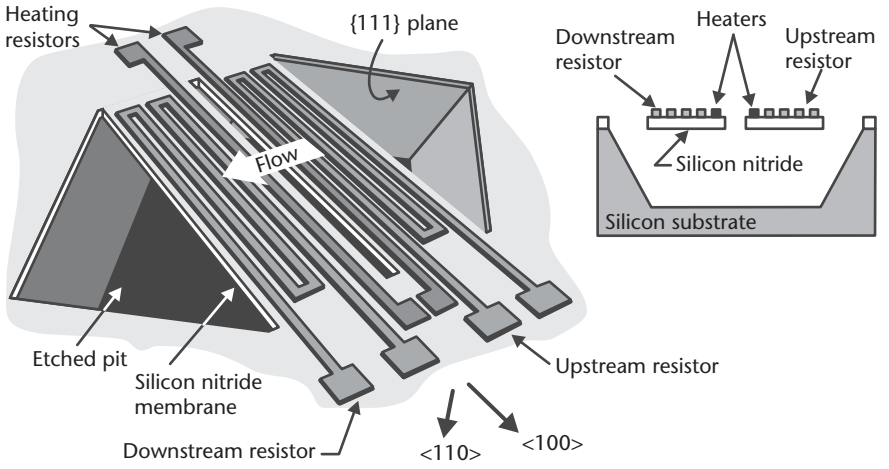


Figure 4.13 Illustration of a micromachined mass flow sensor. Gas flow cools the upstream heater and heats the downstream heater. Temperature-sensitive resistors are used to measure the temperature of each heater and consequently infer the flow rate. The etched pit underneath the heater provides exceptional thermal isolation to the silicon support frame. (*After:* technical sheets on the AWM series of mass airflow sensors, Honeywell, Inc., of Minneapolis, Minnesota, and [16].)

because it can be deposited under low tensile stress, and it retains its structural integrity in most anisotropic etch solutions. The thin-film heaters and sense elements are deposited next by sputtering a thin metal layer (e.g., platinum or nickel) or by the chemical vapor deposition of a heavily doped layer of polysilicon. The thin metal film or polysilicon are then patterned using standard lithography followed by an appropriate etch step. An insulating passivation layer, preferably made of silicon nitride, encapsulates and protects the heating and sense elements. Both silicon nitride layers must then be lithographically patterned in the shape of the two suspended membranes and consequently etched to expose the silicon regions outside of the membrane outline. The final step involves the etching of the silicon in potassium hydroxide or a similar anisotropic etch solution to form the deep cavity. The etch first proceeds in the open silicon regions, and then it progresses underneath the silicon nitride thin film, removing all the silicon and resulting in the suspended silicon nitride membranes. The reason the etch proceeds underneath the silicon nitride layer is because its orientation is in the <100> direction. The etch stops on the {111} crystallographic planes along the periphery of the open silicon areas.

Acceleration Sensors

The first demonstration of a micromachined accelerometer took place in 1979 at Stanford University [17], but it took nearly 15 years before such devices became accepted mainstream products for large-volume applications (see Table 4.3). The overall market for silicon microaccelerometers has been steadily increasing, reaching an estimated $319 million in 2000 [18] and driven primarily by the need for crash sensing in airbag deployment systems. The increase in unit volume has been accompanied by a steady decrease in pricing for automotive applications from an estimated $10 per unit in the early 1990s to less than $2 per unit in 2002. Clearly, low-volume pricing for custom designs remains well above quoted figures for the high-volume automotive markets.

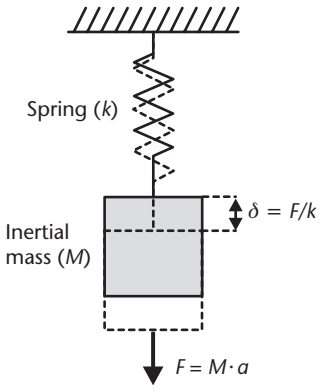
Table 4.3 Some Applications for Micromachined Accelerometers

Measurement	*Application*
Acceleration	Front and side airbag crash sensing
	Electrically controlled car suspension
	Safety belt pretensioning
	Vehicle and traction control systems
	Inertial measurement, object positioning, and navigation
	Human activity for pacemaker control
Vibration	Engine management
	Condition-based maintenance of engines and machinery
	Security devices
	Shock and impact monitoring
	Monitoring of seismic activity
Angles of inclination	Inclinometers and tilt sensing
	Vehicle stability and roll
	Headlight leveling
	Computer peripherals (e.g., joystick, head mounted displays)
	Handwriting recognition (e.g., SmartQuill from British Telecom plc)
	Bridges, ramps, and construction

All accelerometers share a basic structure consisting of an inertial mass suspended from a spring (see Figure 4.14). They differ in the sensing of the relative position of the inertial mass as it displaces under the effect of an externally applied acceleration. A common sensing method is capacitive, in which the mass forms one side of a two-plate capacitor. This approach requires the use of special electronic circuits to detect minute changes in capacitance ($<10^{-15}$ F) and to translate them into an amplified output voltage. Another common method uses piezoresistors to sense the internal stress induced in the spring. In yet a different method, the spring is piezoelectric or contains a piezoelectric thin film, providing a voltage in direct proportion to the displacement. In some rare instances, such as in operation at elevated temperatures, position sensing with an optical fiber becomes necessary. The focus of this section is on capacitive and piezoresistive accelerometers.

The primary specifications of an accelerometer are full-scale range, often given in G, the Earth's gravitational acceleration (1 G = 9.81 m/s^2), sensitivity (V/G), resolution (G), bandwidth (Hz), cross-axis sensitivity, and immunity to shock. The range and bandwidth required vary significantly depending on the application. Accelerometers for airbag crash sensing are rated for a full range of ±50G and a bandwidth of about one kilohertz. By contrast, devices for measuring engine knock or vibration have a range of about 1G, but must resolve small accelerations (<100 μG) over a large bandwidth (>10 kHz). Modern cardiac pacemakers incorporate multiaxis accelerometers to monitor the level of human activity, and correspondingly adjust the stimulation frequency. The ratings on such sensors are ±2G and a bandwidth of less than 50 Hz, but they require extremely low power consumption for battery longevity. Accelerometers for military applications such as fuzing can exceed a rating of 1,000G.

Cross-axis sensitivity assesses the immunity of the sensor to accelerations along directions perpendicular to the main sensing axis. Cross-axis rejection ratios in excess of 40 dB are always desirable. Shock immunity is an important but somewhat subjective specification for the protection of the devices during handling or operation. While one would expect the specification quantified in units of acceleration, it is instead defined in terms of a peculiar but more practical test involving dropping the device from a height of one meter over concrete—the shock impact can easily



Resonant frequency:

$$f_r = \frac{1}{2\pi}\sqrt{\frac{k}{M}}$$

Noise equivalent acceleration:

$$a_{noise} = \sqrt{\frac{8\pi K_B T f_r B}{Q\,M}} \quad ; \quad B < f_r$$

K_B = Boltzmann constant
T = Temperature
B = Bandwidth
Q = Quality factor

Figure 4.14 The basic structure of an accelerometer, consisting of an inertial mass suspended from a spring. The resonant frequency and the noise-equivalent acceleration (due to Brownian noise) are given.

reach a dynamic peak of 10,000G! In addition to achieving a large impact, the drop test excites the various modes of resonance that are liable to cause catastrophic failure.

While many companies offer micromachined acceleration sensor products, a representative set of only four accelerometers follows next, each unique in its design and fabrication.

Piezoresistive Bulk Micromachined Accelerometer

Until only recently, piezoresistive-type acceleration sensors were widely available. Many companies, including GE NovaSensor of Fremont, California, and IC Sensors of San Jose, California (a division of Measurement Specialties, Inc., of Fairfield, New Jersey), offered products using anisotropically etched silicon inertial mass and diffused piezoresistive sense elements. But these products were retired because they could not meet the aggressive pricing requirements of the automotive industry. The product introduction in 1996 by Endevco Corp., of San Juan Capistrano, California, indicates that piezoresistive accelerometers remain in this highly competitive market (see Figure 4.15).

The Endevco sensor consists of three substrates: a lower base; a middle core containing a hinge-like spring, the inertial mass, and the sense elements; and finally a top protective lid [19]. The inertial mass sits inside a frame suspended by the spring. Two thin boron-doped piezoresistive elements in a Wheatstone bridge configuration span the narrow 3.5-μm gap between the outer frame of the middle core and the inertial mass. The piezoresistors are only 0.6 μm thick and 4.2 μm long and are thus very sensitive to minute displacements of the inertial mass. The output in response to an acceleration equal to 1G in magnitude is 25 mV for a Wheatstone bridge excitation of 10V. The thick and narrow hinge structure allows displacement within the plane of

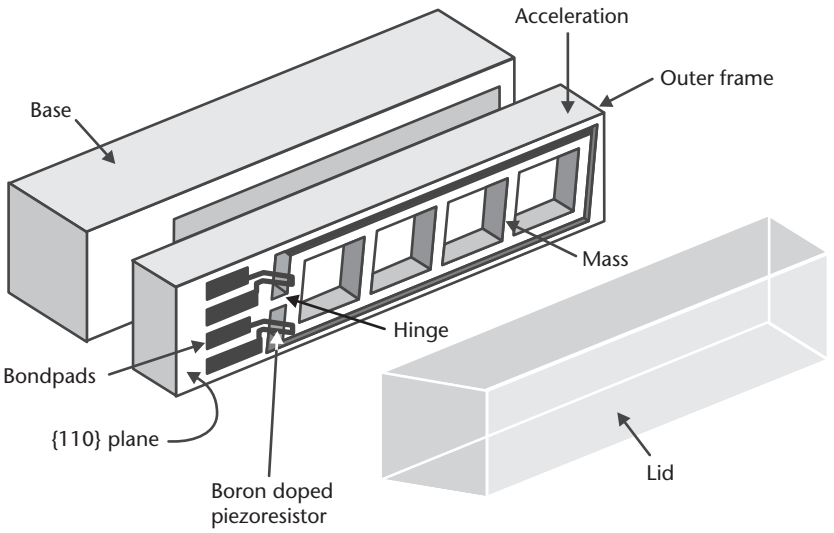


Figure 4.15 Illustration of a piezoresistive accelerometer from Endevco Corp., fabricated using anisotropic etching in a {110} wafer. The middle core contains the inertial mass suspended from a hinge. Two piezoresistive sense elements measure the deflection of the mass. The axis of sensitivity is in the plane of the middle core. The outer frame acts as a stop mechanism to prevent excessive accelerations from damaging the part. (*After:* [19].)

the device, but it is very stiff in directions normal to the wafer, resulting in high immunity to off-axis accelerations. Moreover, the outer frame acts as a stop mechanism that protects the device in the event of excessive acceleration shocks. It takes 6,000G for the inertial mass to touch the frame, and the device can survive shocks in excess of 10,000G. Open apertures reduce the weight of the inertial mass and combine with the stiff hinge to provide a rather high resonant frequency of 28 kHz.

The fabrication process is somewhat unique with its utilization of {110} wafers for the middle core. In this case, the {111} crystallographic planes are orthogonal to the {110} surface of the wafer, which allows the formation of vertical trenches using anisotropic wet etchants. The fabrication begins with boron implantation and diffusion at 1,100°C to form highly doped *p*-type piezoresistors. In order to obtain maximum sensitivity, the piezoresistors are aligned along a <111> direction. A silicon oxide or silicon nitride layer masks the silicon in the form of the inertial mass and hinge during the subsequent anisotropic etch in EDP. The inertial mass is bounded by vertical {111} planes, giving it the shape of a parallelogram whose inside angle is 70.5° (see Chapter 3). Subsequent fabrication steps provide for the deposition and patterning of aluminum electrical contacts and bond pads. Shallow recesses are incorporated in the base and lid substrates before the three-wafer stack is bonded together using low-melting-point glass as the adhesive.

Capacitive Bulk Micromachined Accelerometer

Many companies offer capacitive bulk-micromachined accelerometers. The next example describes the SCA series from VTI Technologies of Vantaa, Finland. The sensor consists of a stack of three bonded silicon wafers, with the hinge spring and inertial mass incorporated in the middle wafer. The inertial mass forms a moveable inner electrode of a variable differential capacitor circuit. The two outer wafers are identical and are simply the fixed electrodes of the two capacitors (see Figure 4.16).

Holes through the inertial mass reduce the damping effect from air trapped in the enclosed cavity, increasing the operating bandwidth of the sensor. Unlike other designs, the contacts to the electrodes are on the side of the die and thus must be defined after the wafer is diced into individual sensor parts. The SCA series of sensors is available in a measuring range from ±0.5G to ±12G. Electronic circuits sense changes in capacitance, then convert them into an output voltage between 0 and 5V. The rated bandwidth is up to 400 Hz for the ±12G accelerometer, the cross-axis sensitivity is less than 5% of output, and the shock immunity is 20,000G.

The particulars of the VTI Technologies process are not publicly available; however, Sasayama et al. [20] describe a process for building a similar part (see Figure 4.17). The three wafers are fabricated separately, then joined at the end by a bonding process, such as anodic bonding, silicon fusion bonding, or even a glass thermocompression bond. The upper and lower wafers are identical and contain a metal electrode patterned with standard lithography over a thin layer of silicon dioxide. The inertial mass and hinge are delineated in the middle wafer using four sequential steps of anisotropic etching in potassium hydroxide or a similar etchant. First, shallow recess cavities are etched on both sides of the wafer. Three distinct masking layers are each deposited and patterned separately. Silicon dioxide and silicon nitride are good material choices. Each of these masking layers is sequentially removed after an etch step in an anisotropic wet etching solution. In essence, the

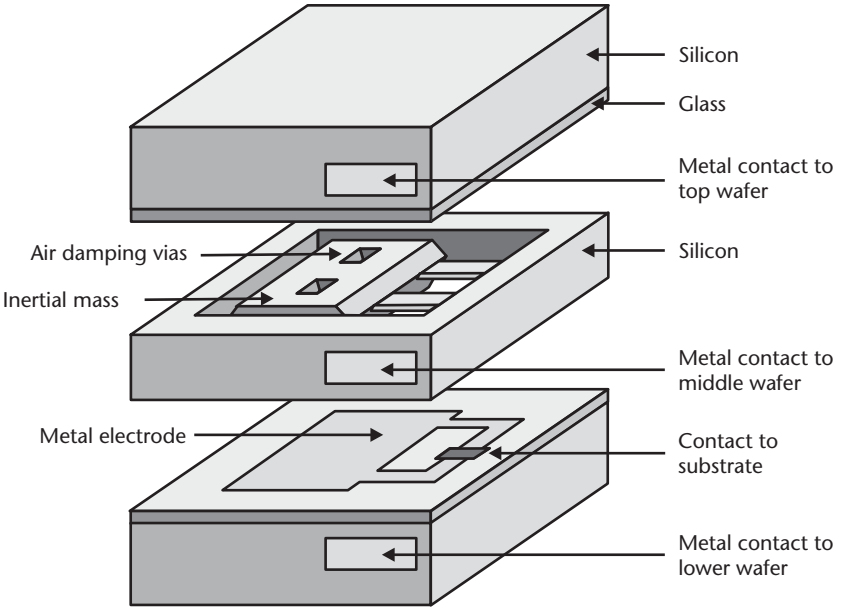


Figure 4.16 Illustration of a bulk micromachined capacitive accelerometer. The inertial mass in the middle wafer forms the moveable electrode of a variable differential capacitive circuit. (*After:* accelerometer product catalog of VTI Technologies of Vantaa, Finland.)

pattern information is encoded in each of the three masking layers. Timed etching simply translates the encoded information into a variable topography in the silicon substrate. The end result is a thin support hinge member with a much thicker inertial mass. The recesses on either side of the mass form the thin gaps for the two-plate sense capacitors.

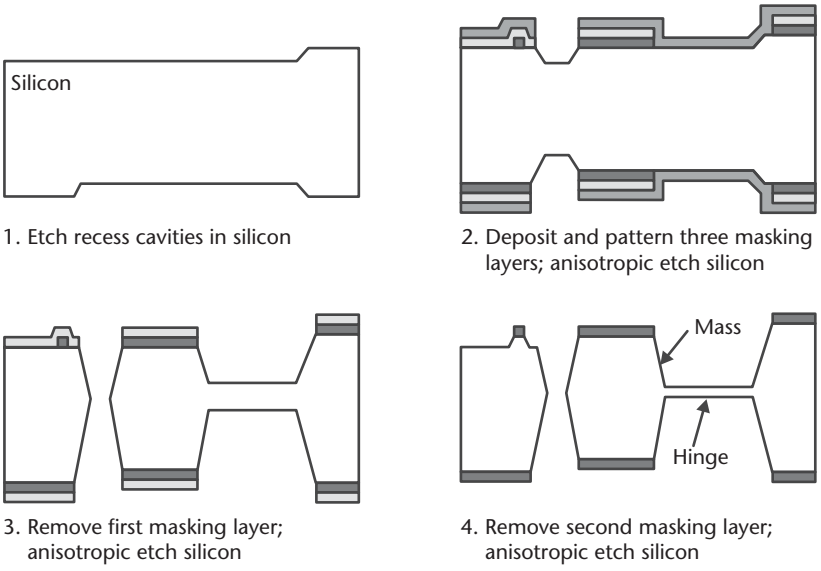


Figure 4.17 Process steps to fabricate the middle wafer containing the hinge and the inertial mass of a bulk micromachined capacitive accelerometer similar to the device from VTI Technologies. (*After:* [20].)

Capacitive Surface Micromachined Accelerometer

Surface micromachining emerged in the late 1980s as a perceived low-cost alternative for accelerometers aimed primarily at automotive applications. Both Robert Bosch GmbH of Stuttgart, Germany, and Analog Devices, Inc., of Norwood, Massachusetts, offer surface micromachined accelerometers, but it is the latter company that benefited from wide publicity to their ADXL product family [21]. The Bosch sensor [22] is incorporated in the Mercedes Benz family of luxury automobiles. The ADXL parts are used on board Ford, General Motors, and other vehicles, as well as inside joysticks for computer games. The surface micromachining fabrication sequence, illustrated in Chapter 3, is fundamentally similar to both sensors, though the Bosch device uses a thicker (10-μm) polysilicon structural element.

Unlike most bulk-micromachined parts, surface-micromachined accelerometers incorporate a suspended comb-like structure whose primary axis of sensitivity lies in the plane of the die. This is often referred to as an *x-axis* (or *y-axis*) type of device, as opposed to *z-axis* sensors where the sense axis is orthogonal to the plane of the die. However, due to the relative thinness of their structural elements, surface micromachined accelerometers suffer from sensitivity to accelerations out of the plane of the die (*z*-axis). Shocks along this direction can cause catastrophic failures.

The ADXL device [21] consists of three sets of 2-μm-thick polysilicon finger-like electrodes (see Figure 4.18). Two sets are anchored to the substrate and are stationary. They form the upper and lower electrode plates of a differential capacitance system, respectively. The third set has the appearance of a two-sided comb whose fingers are interlaced with the fingers of the first two sets. It is suspended approximately 1 μm over the surface by means of two long, folded polysilicon beams acting as suspension springs. It also forms the common middle and displaceable

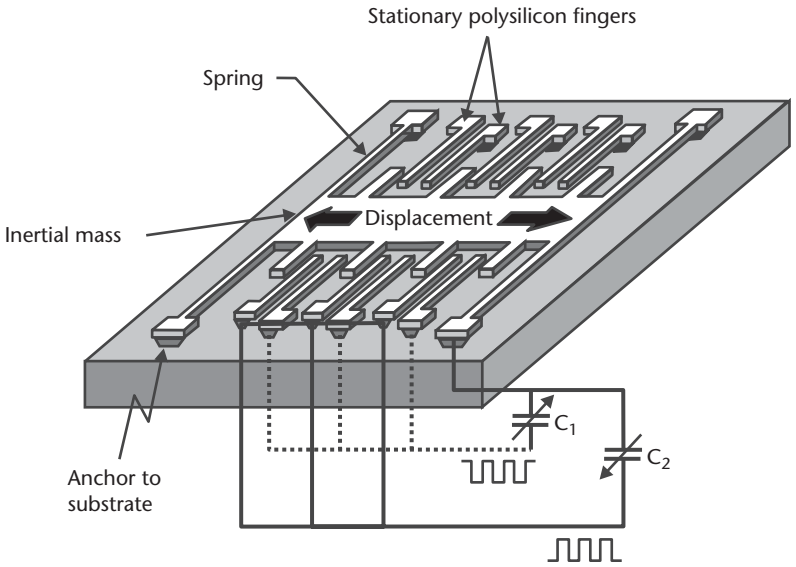


Figure 4.18 Illustration of the basic structure of the ADXL family of surface micromachined accelerometers. A comb-like structure suspended from springs forms the inertial mass. Displacements of the mass are measured capacitively with respect to two sets of stationary finger-like electrodes. (*After:* ADXL data sheets and application notes of Analog Devices, Inc., of Norwood, Massachusetts.)

electrode for the two capacitors. The inertial mass consists of the comb fingers and the central backbone element to which these suspended fingers are attached. Under no externally applied acceleration, the two capacitances are identical. The output signal, proportional to the difference in capacitance, is null. An applied acceleration displaces the suspended structure, resulting in an imbalance in the capacitive half bridge. The differential structure is such that one capacitance increases, and the other decreases. The overall capacitance is small, typically on the order of 100 fF (1 fF = 10^{-15} F). For the ADXL105 (programmable at either ±1G or ±5G), the change in capacitance in response to 1G is minute, about 100 aF (1 aF = 10^{-18} F). This is equivalent to only 625 electrons at an applied bias of one volt and thus must be measured using on-chip integrated electronics to greatly reduce the impact of parasitic capacitance and noise sources, which would be present with off-chip wiring. The basic read-out circuitry consists of a small-amplitude, two-phase oscillator driving both ends of the capacitive half bridge in opposite phases at a frequency of 1 MHz. A capacitance imbalance gives rise to a voltage in the middle node. The signal is then demodulated and amplified. The 1-MHz excitation frequency is sufficiently higher than the mechanical resonant frequency that it produces no actuation force on the plates of the capacitors, provided its dc (average) value is null. The maximum acceleration rating for the ADXL family varies from ±1G (ADXL 105) up to ±100G (ADXL 190). The dynamic range is limited to about 60 dB over the operational bandwidth (typically, 1 to 6 kHz). The small change in capacitance and the relatively small mass combine to give a noise floor that is relatively large when compared to similarly rated bulk micromachined or piezoelectric accelerometers. For the ADXL105, the mass is approximately 0.3 μg, and the corresponding noise floor, dominated by Brownian mechanical noise, is $225\ \mu G/\sqrt{Hz}$. By contrast, the mass for a bulk-micromachined sensor can easily exceed 100 μg.

Applying a large-amplitude voltage at low frequency—below the natural frequency of the sensor—between the two plates of a capacitor gives rise to an electrostatic force that tends to pull the two plates together. This effect enables the application of feedback to the inertial mass: Every time the acceleration pulls the set of suspended fingers away from one of the anchored sets, a voltage significantly larger in amplitude than the sense voltage, but lower in frequency, is applied to the same set of plates, pulling them together and effectively counterbalancing the action of the external acceleration. This feedback voltage is appropriately proportioned to the measured capacitive imbalance in order to maintain the suspended fingers in their initial position, in a pseudostationary state. This electrostatic actuation, also called force balancing, is a form of closed-loop feedback. It minimizes displacement and greatly improves output linearity (because the center element never quite moves by more than a few nanometers). The sense and actuation plates may be the same, provided the two frequency signals (sense and actuation) do not interfere with each other.

A significant advantage to surface micromachining is the ease of integrating two single-axis accelerometers on the same die to form a dual-axis accelerometer, so-called *two-axes*. In a very simple configuration, the two accelerometers are orthogonal to each other. However, the ADXL200 series of dual-axis sensors employs a more sophisticated suspension spring mechanism, where a single inertial mass is shared by both accelerometers.

Capacitive Deep-Etched Micromachined Accelerometer

The DRIE accelerometer developed at GE NovaSensor of Fremont, California, shares its basic comb structure design with the ADXL and Bosch accelerometers. It consists of a set of fingers attached to a central backbone plate, itself suspended by two folded springs (see Figure 4.19). Two sets of stationary fingers attached directly to the substrate complete the capacitive half bridge. The design, however, adds a few improvements. By taking advantage of the third dimension and using structures 50 to 100 μm deep, the sensor gains a larger inertial mass, up to 100 μg, as well as a larger capacitance, up to 5 pF. The relatively large mass reduces mechanical Brownian noise and increases resolution. The high aspect ratio of the spring practically eliminates the sensitivity to z-axis accelerations (out of the plane of the die). Fabrication follows the SFB-DRIE process introduced in Chapter 3.

The sensor, described by van Drieënhuizen et al. [23], uses a 60-μm-thick comb structure for a total capacitance of 3 pF, an inertial mass of 43 μg, a resonant frequency of 3.1 kHz, and an open-loop mechanical sensitivity of 1.6 fF/G. The corresponding mechanical noise is about 10 $\mu G/\sqrt{Hz}$, significantly less than for a surface-micromachined sensor. The read-out circuitry first converts changes in capacitance into frequency. This is accomplished by inserting the two variable capacitors into separate oscillating circuits whose output frequencies are directly proportional to the capacitance. A phase detector compares the two output frequencies and converts the difference into a voltage. The circuit then amplifies the signal before feeding it back to a set of actuation electrodes for force balancing. These electrodes may be distinct from the sense electrodes. Filters set the closed-loop bandwidth to 1 kHz. The overall sensitivity is 700 mV/G for a ±5G device. Early prototypes had a dynamic range of 44 dB limited by electronic 1/f noise in the CMOS circuitry. Recent prototypes with newer implementations of the electronic read-out circuits demonstrated a dynamic range approaching 70 dB over the 1-kHz bandwidth. The SFB-DRIE process is fully compatible with the integration

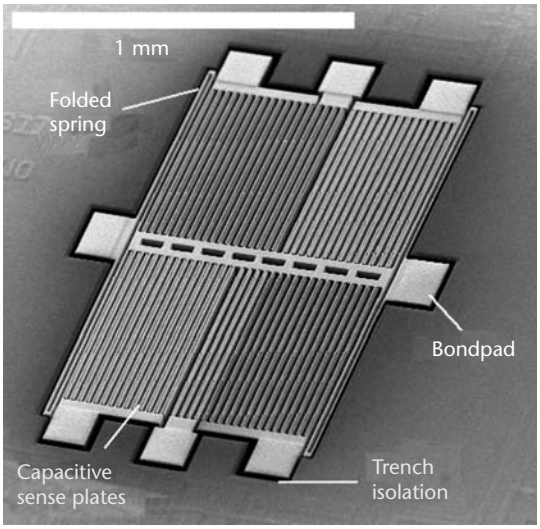


Figure 4.19 Scanning-electron micrograph of a DRIE accelerometer using 60-μm-thick comb structures. (*Courtesy of:* GE NovaSensor of Fremont, California.)

of CMOS circuits next to the mechanical sensing element. The large available capacitance makes the decision to integrate based purely on economics rather than performance.

Angular Rate Sensors and Gyroscopes

Long before the advent of Loran and the satellite-based global positioning system, the gyroscope was a critical navigational instrument used for maintaining a fixed orientation with great accuracy, regardless of Earth rotation. Invented in the nineteenth century, it consisted of a flywheel mounted in gimbal rings. The large angular momentum of the flywheel counteracts externally applied torques and keeps the orientation of the spin axis unaltered. The demonstration of the ring laser gyroscope in 1963 displaced the mechanical gyroscope in many high-precision applications, including aviation. Inertial navigation systems based on ring laser gyroscopes are on board virtually all commercial aircraft. Gyroscopes capable of precise measurement of rotation are very expensive instruments, costing many thousands of dollars. An article published in 1984 by the IEEE reviews many of the basic technologies for gyroscopes [24].

The gyroscope derives its precision from the large angular momentum that is proportional to the heavy mass of the flywheel, its substantial size, and its high rate of spin (see Figure 4.20). This, in itself, precludes the use of miniature devices for useful gyroscopic action; the angular momentum of a miniature flywheel is miniscule. Instead, micromachined sensors that detect angular rotation utilize the Coriolis effect. Fundamentally, such devices are strictly angular-rate or yaw-rate sensors, measuring angular velocity. However, they are colloquially but incorrectly referred to as gyroscopes.

The Coriolis effect, named after the French physicist Gaspard Coriolis, manifests itself in numerous weather phenomena, including hurricanes and tornadoes, and is a direct consequence of a body's motion in a rotating frame of reference

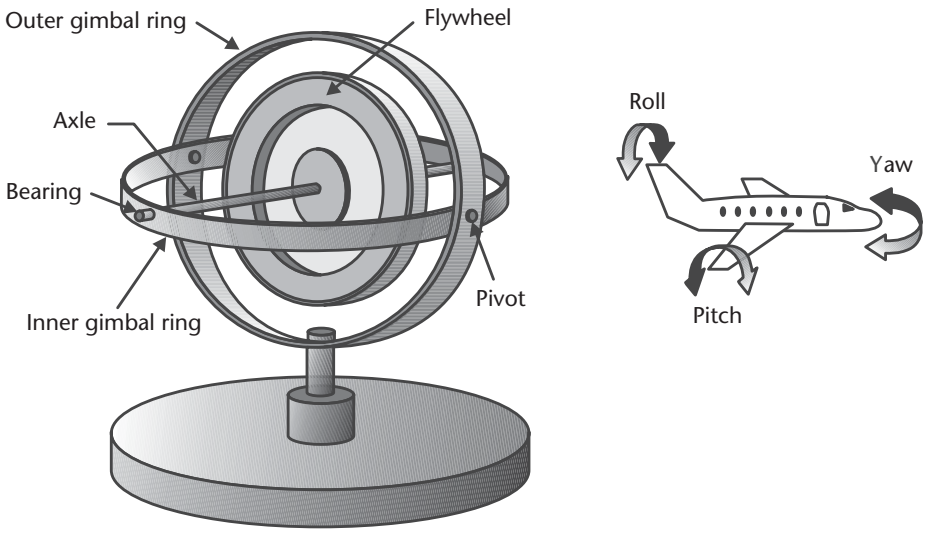


Figure 4.20 Illustration of a conventional mechanical gyroscope and the three rotational degrees of freedom it can measure.

(see Figure 4.21). To understand it, let us imagine an automobile driving from Seattle, Washington (lat. 48° N), to Los Angeles, California (lat. 34° N). At the beginning of its journey, the car in Seattle is actually moving eastward with the rotation of Earth (the rotating frame of reference) at about 1120 km/h[1]. At the end of its journey in Los Angeles, its eastward velocity is 1,385 km/h. As the car moves south across latitudes, its eastward velocity must increase from 1,120 to 1,385 km/h; otherwise, it will continuously slip and never reach its destination. The road—effectively the rotating surface—imparts an eastward acceleration to maintain the vehicle on its course. This is the Coriolis acceleration. In general, the Coriolis acceleration is the acceleration that must be applied in order to maintain the heading of a body moving on a rotating surface [25].

All micromachined angular rate sensors have a vibrating element at their core—this is the moving body. In a fixed frame of reference, a point on this element oscillates with a velocity vector **v**. If the frame of reference begins to rotate at a rate Ω, this point is then subject to a Coriolis force and a corresponding acceleration equal to $2\Omega \times \mathbf{v}$ [26]. The vector cross operation implies that the Coriolis acceleration and the resulting displacement at that point are perpendicular to the oscillation. This, in effect, sets up an energy transfer process from a primary mode of oscillation into a secondary mode that can be measured. It is this excitation of a secondary resonance mode that forms the basis of detection using the Coriolis effect. In beam structures, these two frequencies are distinct with orthogonal displacements. But for highly symmetrical elements, such as rings, cylinders, or disks, the resonant frequency is degenerate, meaning there are two distinct modes of resonance sharing the same oscillation frequency. This degeneracy causes the temporal excitation signal (primary mode) to be in phase quadrature with the sense signal (secondary mode), thus minimizing coupling between these two modes and improving sensitivity and

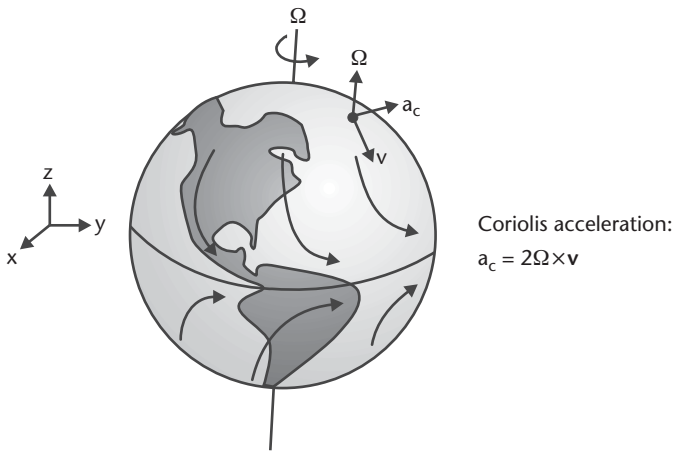


Figure 4.21 Illustration of the Coriolis acceleration on an object moving with a velocity vector **v** on the surface of Earth from either pole towards the equator. The Coriolis acceleration deflects the object in a counterclockwise manner in the northern hemisphere and a clockwise direction in the southern hemisphere. The vector Ω represents the rotation of the planet.

1. The velocity at the equator is 1,670 km/h. The velocity at latitude 48° N is 1,670 km/h multiplied by cos 48°.

accuracy [27]. Additionally, the degeneracy tends to minimize the device's sensitivity to thermal errors, aging, and long-term frequency drifts.

A simple and common implementation is the tuning-fork structure (see Figure 4.22). The two tines of the fork normally vibrate in opposite directions in the plane of the fork (flexural mode). The Coriolis acceleration subjects the tips to a displacement perpendicular to the primary mode of oscillation, forcing each tip to describe an elliptical path. Rotation, hence, excites a secondary vibration torsional mode around the stem with energy transferred from the primary flexural vibration of the tines. Quartz tuning forks such as those from BEI Technologies, Systron Donner Inertial Division of Concord, California, use the piezoelectric properties of the material to excite and sense both vibration modes. The tuning-fork structure is also at the core of a micromachined silicon sensor from Daimler Benz AG that will be described later. Other implementations of angular rate sensors include simple resonant beams, vibrating ring shells, and tethered accelerometers, but all of them exploit the principle of transferring energy from a primary to a secondary mode of resonance. Of all the vibrating angular-rate structures, the ring shell or cylinder is the most promising for inertial and navigational-grade performance because of the frequency degeneracy of its two resonant modes.

The main specifications of an angular-rate sensor are full-scale range (expressed in °/s or °/hr; scale factor or sensitivity [V/(°/s)]; noise, also known as angle random walk [$^\circ/(s\cdot\sqrt{Hz})$]; bandwidth (Hz); resolution (°/s); and dynamic range (dB), the latter two being functions of noise and bandwidth. Short- and long-term drift of the output, known as bias drift, is another important specification (expressed in °/s or °/hr). As is the case for most sensors, angular-rate sensors must withstand shocks of at least 1,000G.

Micromachined angular-rate sensors have largely been unable to deliver a performance better than rate grade. These are devices with a dynamic range of only 40 dB, a noise figure larger than $0.1^\circ/\left(s\cdot\sqrt{Hz}\right)$, and a bias drift worse than 10 °/hr. By comparison, inertial grade sensors and true gyroscopes deliver a dynamic range of over 100 dB, a noise less than $0.001^\circ/\left(hr\cdot\sqrt{Hz}\right)$, and a bias drift better than 0.01 °/hr [28]. The advantage of micromachined angular-rate sensors lies in their

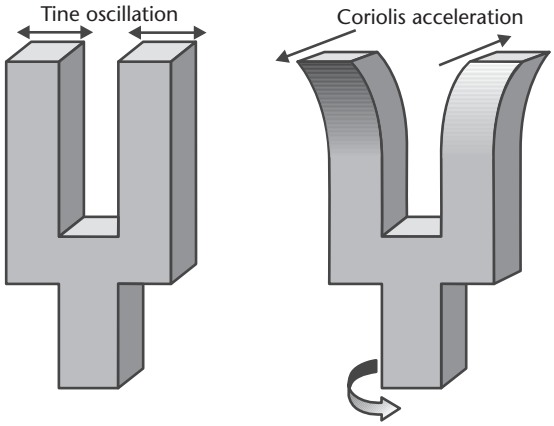


Figure 4.22 Illustration of the tuning-fork structure for angular-rate sensing. The Coriolis effect transfers energy from a primary flexural mode to a secondary torsional mode.

small size and low cost, currently less than $10. They are slowly gaining acceptance in automotive applications, in particular, for vehicle stability systems. The sensor detects any undesired yaw of a vehicle due to poor road conditions and feeds the information to a control system, which may activate the antilock braking system (ABS) or the traction control system (TCS) to correct the situation. The Mercedes Benz ML series of sport utility vehicles incorporates a silicon angular-rate sensor from Robert Bosch GmbH for vehicle stability.

The selection of commercially available micromachined yaw-rate sensors remains limited, but many manufacturers have publicly acknowledged the existence of development programs. The sensors from Delphi Delco Electronics Systems, Robert Bosch GmbH, Daimler Benz AG, and Silicon Sensing Systems illustrate four vibratory-type angular-rate sensors distinct in their structure as well as excitation and sense methods.

Micromachined Angular-Rate Sensor from Delphi Delco Electronics Systems

The sensor from Delphi Delco Electronics Systems of Kokomo, Indiana [29], a division of Delphi Corporation of Troy, Michigan, includes at its core a vibrating ring shell based on the principle of the ringing wine glass discovered in 1890 by G. H. Bryan. He observed that the standing-wave pattern of the wine glass did not remain stationary in inertial space but participated in the motion as the glass rotated about its stem.

The complete theory of vibrating-ring angular-rate sensors is well developed [30]. The ring shell, anchored at its center to the substrate, deforms as it vibrates through a full cycle from a circle to an ellipse, back to a circle, then to an ellipse rotated at right angles to the first ellipse, then back to the original circle (see Figure 4.23). The points on the shell that remain stationary are called nodes, whereas the points that undergo maximal deflection are called antinodes. The nodes and antinodes form a vibration pattern—or standing-wave pattern—around the ring. The pattern is characteristic of the resonance mode. Because of symmetry, a ring shell possesses two frequency-degenerate resonant modes with their vibration patterns offset by 45° with respect to each other. Hence, the nodes of the first mode coincide with the antinodes of the second mode. The external control electronics excite only one of the two modes—the primary mode. But under rotation, the Coriolis effect excites the second resonance mode, and energy transfer occurs between the two modes. Consequently, the deflection amplitude builds up at the antinodes of the second mode—also, the nodes of the first mode. The overall vibration becomes a linear combination of the two modes with a new set of nodes and antinodes forming a vibration pattern rotated with respect to the pattern of the primary mode. It is this lag that Bryan heard in his spinning wine glass. In an open-loop configuration, the deflection amplitude at the nodes and antinodes is a measure of the angular rate of rotation. Alternatively, the angular shift of the vibration pattern is another measure. In a closed-loop configuration, electrostatic actuation by a feedback voltage applied to the excitation electrodes nulls the secondary mode and maintains a stationary vibration pattern. The angular rate becomes directly proportional to this feedback voltage.

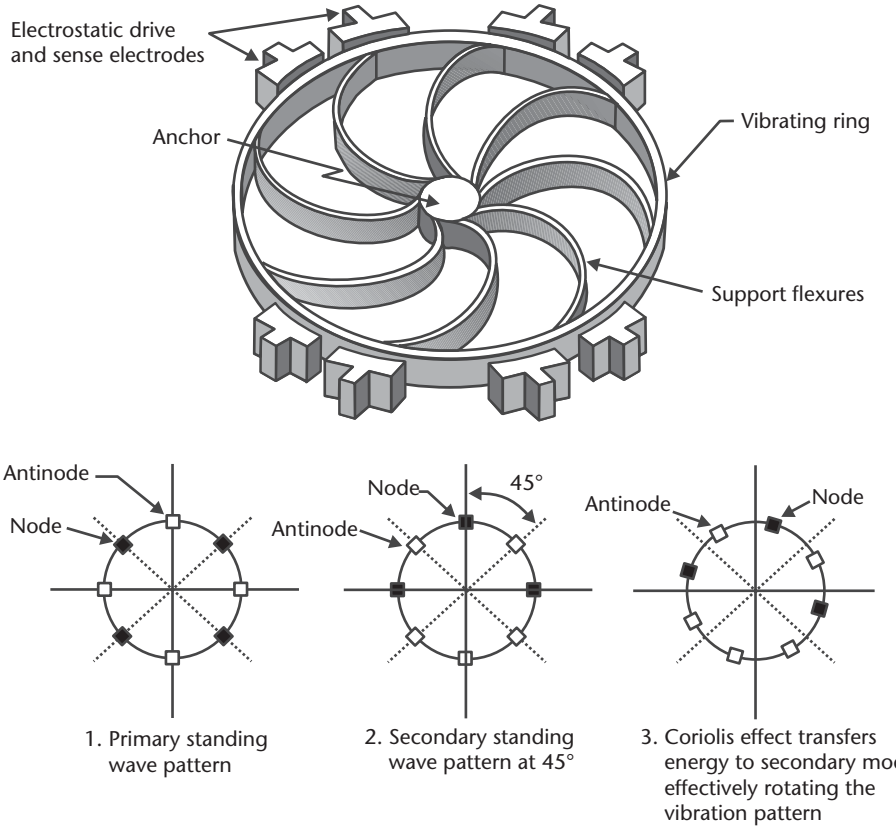


Figure 4.23 Illustration of the Delphi Delco angular-rate sensor and the corresponding standing-wave pattern. The basic structure consists of a ring shell suspended from an anchor by support flexures. A total of 32 electrodes (only a few are shown) distributed around the entire perimeter of the ring excite a primary mode of resonance using electrostatic actuation. A second set of distributed electrodes capacitively sense the vibration modes. The angular shift of the standing-wave pattern is a measure of the angular velocity. (*After:* [29].)

A total of 32 electrodes positioned around the suspended ring shell provide the electrostatic excitation drive and sense functions. Of this set, eight electrodes strategically positioned at 45° intervals—at the nodes and antinodes—capacitively sense the deformation of the ring shell. Appropriate electronic circuits complete the system control functions, including feedback. A phased-locked loop (PLL) drives the ring into resonance through the electrostatic drive electrodes and maintains a lock on the frequency. Feedback is useful to electronically compensate for the mechanical poles and increase the closed-loop bandwidth of the sensor. Additionally, a high mechanical quality factor increases the closed-loop system gain and sensitivity.

The fabrication process is similar to the electroplating and molding process described in Chapter 3, except that the substrate includes preprocessed CMOS control circuitry. The mold is made of photoresist, and the electroplated nickel ring shell is 15 to 50 μm thick. Finally, packaging is completed in vacuum in order to minimize air damping of the resonant ring and provide a large quality factor. Researchers at the University of Michigan demonstrated a polysilicon version of the sensor with improved overall performance.

The demonstrated specifications of the Delphi Delco sensor over the temperature range of –40° to +125°C include a resolution of 0.5°/s over a bandwidth of 25 Hz, limited by noise in the electronic circuitry. The nonlinearity in a rate range of ±100 °/s is less than 0.2°/s. The sensor survives the standard automotive shock test: a drop from a height of one meter. The specifications are adequate for most automotive and consumer applications.

Angular-Rate Sensor from Silicon Sensing Systems

The CRS family of yaw-rate sensors from Silicon Sensing Systems, a joint venture between BAE Systems of Plymouth, Devon, England, and Sumitomo Precision Products Company of Japan, is aimed at commercial and automotive applications. It also uses a vibratory ring shell similar to the sensor from Delphi Delco but differs on the excitation and sense methods. Electric current loops in a magnetic field, instead of electrostatic electrodes, excite the primary mode of resonance. These same loops provide the sense signal to detect the angular position of the vibration pattern (see Figure 4.24).

The ring, 6 mm in diameter, is suspended by eight flexural beams anchored to a 10-mm-square frame. Eight equivalent current loops span every two adjacent support beams. A current loop starts at a bond pad on the frame, traces a support beam to the ring, continues on the ring for one eighth of the circumference, then moves onto the next adjacent support beam, before ending on a second bond pad. Under this scheme, each support beam carries two conductors. A Samarium-Cobalt permanent magnet mounted inside the package provides a magnetic field perpendicular to the beams. Electromagnetic interaction between current in a loop and the magnetic

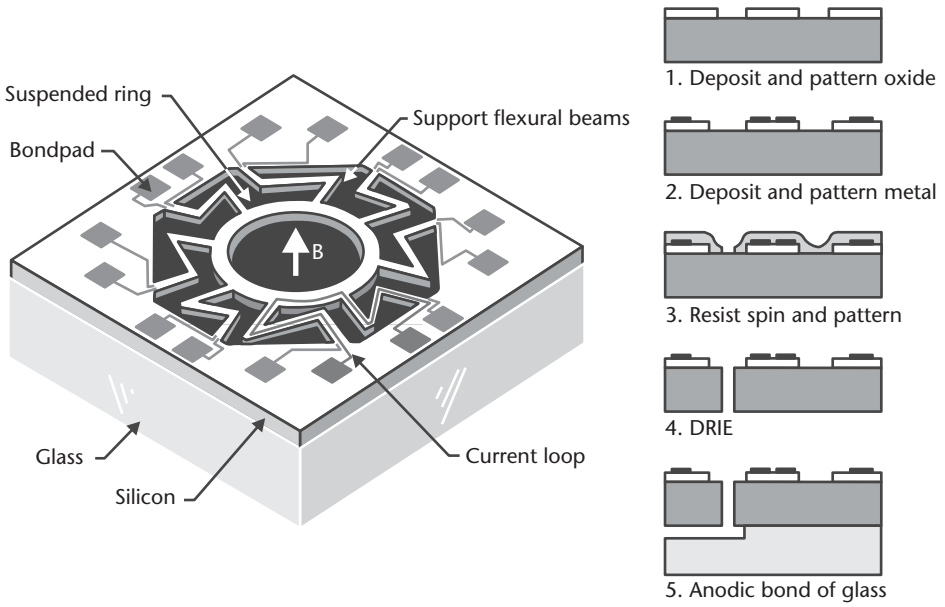


Figure 4.24 Illustration of the CRS angular-rate sensor from Silicon Sensing Systems and corresponding fabrication process. The device uses a vibratory ring shell design, similar to the Delphi Delco sensor. Eight current loops in a magnetic field, *B,* provide the excitation and sense functions. For simplicity, only one of the current loops is shown. (*After:* product data sheet of Silicon Sensing Systems.)

field induces a Lorentz force. Its radial component is responsible for the oscillation of the ring in the plane of the die at approximately 14.5 kHz—the mechanical resonant frequency of the ring. The sensing mechanism measures the voltage induced around one or more loops in accordance with Faraday's law: As the ring oscillates, the area of the current loop in the magnetic flux changes, generating a voltage. Two diametrically opposite loops perform a differential voltage measurement. One can simplistically view an actuating and a sensing loop as the primary and secondary windings of a transformer; the electromagnetic coupling between them depends on the ring vibration pattern and thus on the angular rate of rotation.

Closed-loop feedback improves the overall performance by increasing the bandwidth and reducing the system's sensitivity to physical errors. Two separate feedback loops with automatic gain control circuits maintain a constant oscillation amplitude for the primary mode of resonance and a zero amplitude for the secondary resonance mode. The feedback voltage required to null the secondary mode is a direct measure of the rate of rotation.

The fabrication of the sensor is relatively simple. A silicon dioxide layer is deposited on a silicon wafer, then lithographically patterned and etched. The silicon dioxide layer serves to electrically isolate the current loops. A metal layer is sputter deposited, patterned, and etched to define the current loops as well as the bond pads. A layer of photoresist is spun on and patterned in the shape of the ring and support flexural beams. The photoresist then serves as a mask for a subsequent DRIE step to etch trenches through the wafer. After removal of the photoresist mask, the silicon wafer is anodically bonded to a glass wafer with a previously defined shallow cavity on its surface. Little is available in the open literature on the packaging, but it is clear from the need to include a permanent magnet that the packaging is custom and specific to this application.

The specification sheet of the CRS03-02 gives an output scale factor of 20 mV/(°/s) with a variation of ±3% over a temperature range from –40° to +85°C. The noise is less than 1 mV rms from 3 to 10 Hz. The nonlinearity in a rate range of ±100 °/s is less than 0.5 °/s. The operating current is a relatively large 50 mA at a nominal 5-V supply.

Angular-Rate Sensor from Daimler Benz

The sensor from Daimler Benz AG of Stuttgart, Germany [31], is a strict implementation of a tuning fork using micromachining technology (see Figure 4.25). The tines of the silicon tuning fork vibrate out of the plane of the die, driven by a thin-film piezoelectric aluminum nitride actuator on top of one of the tines. The Coriolis forces on the tines produce a torquing moment around the stem of the tuning fork, giving rise to shear stresses that can be sensed with diffused piezoresistive elements. The shear stress is maximal on the center line of the stem and corresponds with the optimal location for the piezoresistive sense elements.

The high precision of micromachining is not sufficient to ensure the balancing of the two tines and the tuning of the two resonant frequencies—recall from the earlier discussion that the vibration modes of a tuning fork are not degenerate. An imbalance in the tines produces undesirable coupling between the excitation and sense resonant modes, which degrades the resolution of the device. A laser ablation step precisely removes tine material and provides calibration of the tuning fork. For this

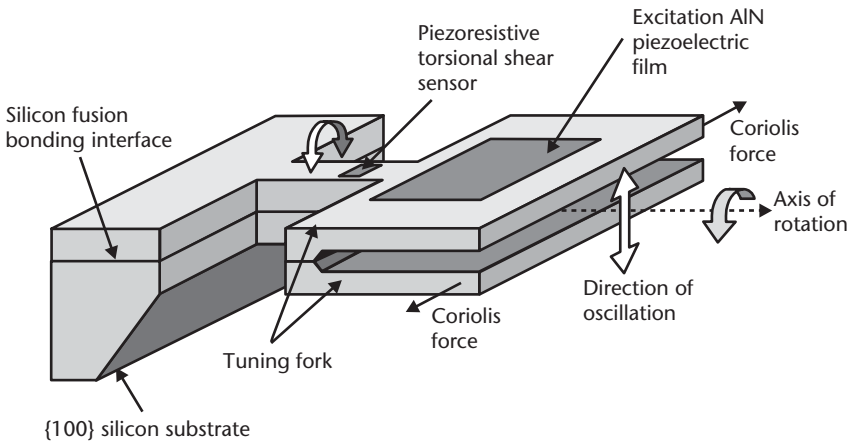


Figure 4.25 Illustration of the angular-rate sensor from Daimler Benz. The structure is a strict implementation of a tuning fork in silicon. A piezoelectric actuator excites the fork into resonance. The Coriolis force results in torsional shear stress in the stem, which is measured by a piezoresistive sense element. (*After:* [31].)

particular design, all resonant modes of the fork are at frequencies above 10 kHz. To minimize coupling to higher orders, the primary and secondary modes are separated by at least 10 kHz from all other remaining modes. The choice of crystalline silicon for tine material allows achieving a high quality factor (~ 7,000) at pressures below 10^{-5} bar.

The fabrication process is distinct from that of other yaw-rate sensors in its usage of SOI substrates (see Figure 4.26). The crystalline silicon over the silicon dioxide layer defines the tines. The thickness control of the tines is accomplished at the beginning of the process by the precise epitaxial growth of silicon over the SOI substrate. The thickness of the silicon layer, and consequently of the tine, varies between 20 and 200 μm, depending on the desired performance of the sensor. Lithography followed by a shallow silicon etch in tetramethyl ammonium hydroxide (TMAH) define 2-μm-deep cavities in two mirror-image SOI substrates. Silicon

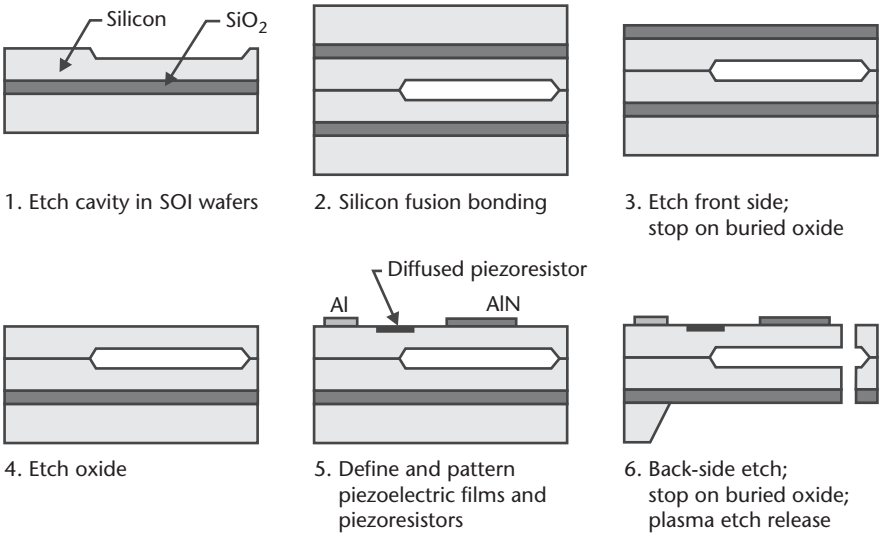


Figure 4.26 The main fabrication steps for the Daimler Benz micromachined angular-rate sensor.

fusion bonding brings these substrates together such that the cavities are facing each other. The cavity depth determines the separation between the two tines. An etch step in TMAH removes the silicon on the front side and stops on the buried silicon dioxide layer which is subsequently removed in hydrofluoric acid. The following steps define the piezoelectric and piezoresistive elements on the silicon surface. Diffused piezoresistors are formed using ion implantation and diffusion. Piezoelectric aluminum nitride is then deposited by sputtering aluminum in a controlled nitrogen and argon atmosphere. This layer is lithographically patterned and etched in the shape of the excitation plate over the tine. Aluminum is then sputtered and patterned to form electrical interconnects and bond pads. Finally, a TMAH etch step from the back side removes the silicon from underneath the tines. The buried silicon dioxide layer acts as an etch stop. An anisotropic plasma etch from the front side releases the tines.

The measured frequency of the primary, flexural mode (excitation mode) was 32.2 kHz, whereas the torsional secondary mode (sense mode) was 245 Hz lower. Typical of tuning forks, the frequencies exhibited a temperature dependence. For this particular technology, the temperature coefficient of frequency is –0.85 Hz/°C.

Angular-Rate Sensor from Robert Bosch

This sensor from Robert Bosch GmbH of Stuttgart, Germany, is unique in its implementation of a mechanical resonant structure equivalent to a tuning fork [32]. An oscillator system consists of two identical masses coupled to each other by a spring and suspended from an outer frame by two other springs (see Figure 4.27). Such a

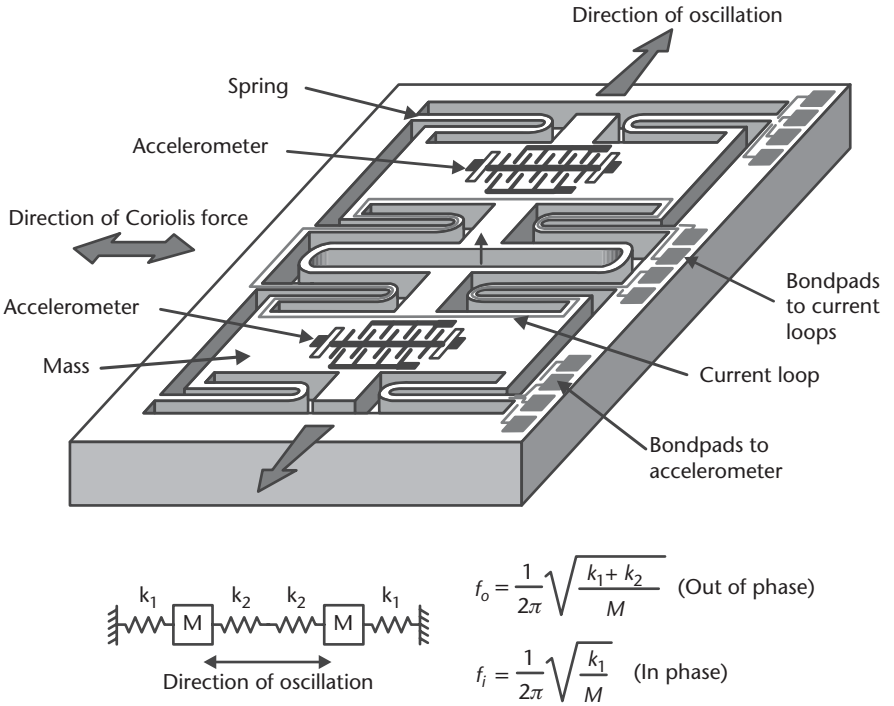


Figure 4.27 Illustration of the yaw-rate sensor from Robert Bosch GmbH. A simple mechanical model shows the two masses and coupling springs. (*After:* [32].)

coupled system has two resonant frequencies: in phase, and out of phase. In the in-phase oscillation mode, the instantaneous displacements of the two masses are in the same direction. In the out-of-phase mode, the masses are moving, at any instant, in opposite directions. A careful selection of the coupling spring provides sufficient separation between the in-phase and out-of-phase resonant frequencies. Lorentz forces generated by an electric current loop within a permanent magnetic field excite only the out-of-phase mode. The oscillation electromagnetically induces a voltage in a second current loop that provides a feedback signal proportional to the velocity of the masses. The resulting Coriolis forces on the two masses are in opposite directions but orthogonal to the direction of oscillation. Two polysilicon surface-micromachined accelerometers with capacitive comb structures (similar in their basic operation to the ADXL family of sensors) measure the Coriolis accelerations for each of the masses. The difference between the two accelerations is a direct measure of the angular yaw rate, whereas their sum is proportional to the linear acceleration along the accelerometer's sensitive axis. Electronic circuits perform the addition and subtraction functions to filter out the linear acceleration signal.

For the Bosch sensor, the out-of-phase resonant frequency is 2 kHz, and the maximum oscillation amplitude at this frequency is 50 μm. The measured quality factor of the oscillator at atmospheric pressure is 1,200, sufficiently large to excite resonance with small Lorentz forces. The stimulated oscillation subjects the masses to large accelerations reaching approximately 800G. Though they are theoretically perpendicular to the sensitive axis of the accelerometers, in practice, some coupling remains, which threatens the signal integrity. However, because the two temporal signals are in phase quadrature, adopting synchronous demodulation methods allows the circuits to filter the spurious coupled signal with a rejection ratio exceeding 78 dB. This is indeed a large rejection ratio but insufficient to meet the requirements of inertial navigation.

The peak Coriolis acceleration for a yaw rate of 100°/s is only 200 mG. This requires extremely sensitive accelerometers with compliant springs. The small Coriolis acceleration further emphasizes the need for perfect orthogonality between the sense and excitation axes. Closed-loop position feedback of the acceleration sense element compensates for the mechanical poles and increases the bandwidth of the accelerometers to over 10 kHz.

The fabrication process simultaneously encompasses bulk and surface micromachining: the former to define the masses and the latter to form the comb-like accelerometers (see Figure 4.28). The process sequence begins by depositing a 2.5-μm layer of silicon dioxide on a silicon substrate. Epitaxy over the oxide layer grows a 12-μm-thick layer of heavily doped *n*-type polysilicon. This layer forms the basis for the surface-micromachined sensors and is polycrystalline because of the lack of a seed crystal during epitaxial growth. In the next step, aluminum is deposited by sputtering and patterned to form electrical interconnects and bond pads. Timed etching from the back side using potassium hydroxide thins the central portion of the wafer to 50 μm. Two sequential DRIE steps define the structural elements of the accelerometers and the oscillating masses. The following step involves etching the sacrificial silicon dioxide layer using a gas phase process (e.g., hydrofluoric acid vapor) to release the polysilicon comb structures. Finally, a

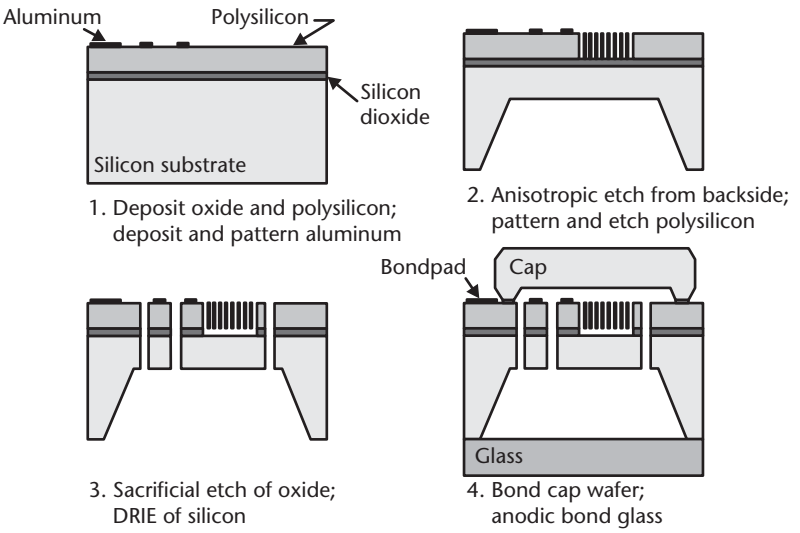


Figure 4.28 Illustration of the fabrication process for the yaw-rate sensor from Robert Bosch GmbH. (*After:* [32].)

protective silicon cap wafer that contains a recess cavity is bonded on the front side using a low temperature seal glass process. A glass wafer anodically bonded to the back side seals the device. The final assembly brings together the silicon sensor and the electronic circuits inside a metal can whose cover holds a permanent magnet.

The sensitivity of the device is 18 mV/(°/s) in the range of ±100 °/s over –40° to +85°C. The temperature dependence of the uncompensated sensor causes an offset amplitude of 0.5 °/s over the specified temperature range, but signal conditioning circuits reduce this dependence by implementing appropriate electronic temperature compensation schemes.

Carbon Monoxide Gas Sensor

Many gas sensors operate on the principle of modulating the resistance of a metal-oxide element by adsorption of gas molecules to its surface. The adsorbed gas molecules interact with the surface of such a wide-bandgap semiconductor to trap one or more conduction electrons, effectively reducing the surface conductivity. The resistance is inversely proportional to a fractional power of the gas concentration. The class of sensor materials include the oxides of tin (SnO_2), titanium (TiO_2), indium (In_2O_3), zinc (ZnO), tungsten (WO_3), and iron (Fe_2O_3). Each metal oxide is sensitive to different gases. For example, tin oxide is effective at detecting alcohol, hydrogen, oxygen, hydrogen sulfide, and carbon monoxide. Indium oxide, by contrast, is sensitive to ozone (O_3); zinc oxide is useful for detecting halogenated hydrocarbons. Unfortunately, most are adversely affected by humidity, which must be controlled at all times. In addition, variations in material properties require that each sensor is individually calibrated.

The MiCS series of carbon monoxide sensors from MicroChemical Systems SA of Switzerland [33] is based on an earlier implementation by Motorola that incorporated a tin-oxide, thin-film sense resistor over a polysilicon resistive heater [34]. The role of the heater is to maintain the sensor at an operating temperature between 100°

and 450°C, thus reducing the deleterious effects of humidity. The sense resistor and the heater reside over a 2-μm-thick silicon membrane to minimize heat loss through the substrate. Consequently, a mere 47 mW is sufficient to maintain the membrane at 400°C. There are a total of four electrical contacts: two connect to the tin-oxide resistor, and the other two connect to the polysilicon heater. The simplest method to measure resistance is to flow a constant current through the sense element and record the output voltage (see Figure 4.29).

The particulars of the fabrication process for the MiCS carbon monoxide sensor and its predecessor by Motorola are not publicly disclosed, but demonstrations of similar devices exist in the literature. A simple process would begin with the forming of a heavily doped, *p*-type, 2-μm-thick layer of silicon either by epitaxial growth or, alternatively, by ion implantation and annealing. The deposition of a silicon nitride layer follows. A chemical vapor deposition (CVD) step provides a polysilicon film that is later patterned and etched in the shape of the heater. The polysilicon film is doped either *in situ* during the CVD process or by ion implantation and subsequent annealing. An oxide layer is then deposited and contact holes etched in it. The purpose of this layer is to electrically isolate the polysilicon heater from the tin-oxide sense element. The tin-oxide layer is deposited by sputtering tin and oxidizing it at approximately 400°C. An alternative deposition process is sol-gel, starting with a tin-based organic precursor and curing by firing at an elevated temperature. The tin-oxide layer is patterned using standard lithography and etched in the shape of the sense element. Sputtered and patterned aluminum provides contact metallization. Finally, an etch from the back side in potassium hydroxide or EDP forms a thin membrane by stopping on the heavily doped *p*-type surface silicon layer. Naturally, a masking layer (e.g., silicon nitride) on the back side of the substrate and protection of the front side are necessary. It is also possible to etch all of the silicon and stop at

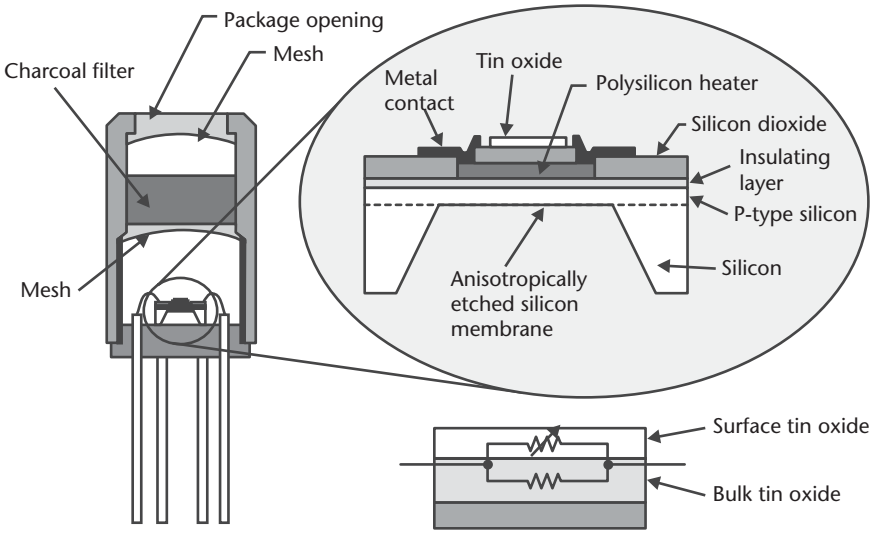


Figure 4.29 Illustration of a carbon monoxide sensor, its equivalent circuit model, and the final packaged part. The surface resistance of tin-oxide changes in response to carbon monoxide. A polysilicon heater maintains the sensor at a temperature between 100° and 450°C in order to reduce the adverse effects of humidity. (*After:* [34].)

the silicon nitride layer to further increase thermal isolation and improve the sensor's performance.

The operation of the earlier sensor by Motorola consists of applying to the heater a 5-V pulse for 5s, followed by a 1-V pulse lasting 10s. The corresponding temperature is 400°C during the first interval, decreasing to 80°C during the second pulse. To maintain consistency, the resistance measurement always occurs at the same time during the interval—in this case, at 9.5s into the second 10-s long pulse. The MiCS sensor demonstrates a response from 10 to 1,000 parts per million (ppm) of carbon monoxide (CO) over a humidity range of 5 to 95%. The output signal shows a square-root dependence on CO concentration, with little dependence on humidity for CO concentrations above 60 ppm.

Actuators and Actuated Microsystems

The physical world is not still but is rather very dynamic and full of motion. If sensors extend our faculties of sight, hearing, smell, and touch, then actuators must be the extension to our hands and fingers. They give us the agility and dexterity to manipulate physical parameters well beyond our reach. It is not surprising that the promise to control at a miniature scale is fascinating. Wouldn't the surgeon dream of electronically controlled precision surgical tools? And what to do when our sensors tell us of a need to locally act and control on a microscopic scale? It is actuation that affords us the ability to apply this type of feedback.

In this section, we address the use of micromachined actuators primarily in industrial and automotive applications—though it is well understood that with minor modifications, these actuators can be applied to other markets. Inkjet heads and microvalves are perhaps the most notable examples to discuss. Micromachined pumps are also emerging as new products of the future. The complexity of actuated microsystems continues to increase as the technology matures, accompanied with a rapidly rising level of integration. For instance, novel microfluidic systems now integrate valves and pumps, as well as various types of sensors and interconnecting channels, and they have become a separate field of study and development [35].

Thermal Inkjet Heads

The thermal inkjet print head, ubiquitous in today's printers for personal computers, receives frequent mention as a premier success story of MEMS technology. While thermal inkjet technology is a commercial success for Hewlett-Packard, Inc., of Palo Alto, California, and a few other companies, there is little in it that originates from silicon MEMS *per se*. Early generations of inkjet heads used electroformed nickel nozzles [10, 36, 37]. More recent models use nozzle plates drilled by laser ablation [38]. Silicon micromachining is not likely to compete with these traditional technologies on a cost basis. However, applications that require high-resolution printing will probably benefit from micromachined nozzles. At a resolution of 1,200 dots per inch (dpi), the spacing between adjacent nozzles in a linear array is a mere 21 μm. A greater number of laser-drilled nozzles on a head raises the cost, while the cost remains constant as holes are added using batch-fabrication methods. Nonetheless, the nozzles continue to be made in nickel plates, but micromachining technology is now necessary to integrate a large number of

microheaters on a silicon chip. High-performance inkjet technology represents an excellent illustration of how micromachining has become a critical and enabling element in a more complex system.

The device from Hewlett-Packard illustrates the basic principle of thermal inkjet printing (see Figure 4.30) [38, 39]. A well under an orifice contains a small volume of ink held in place by surface tension. To fire a droplet, a thin-film resistor made of tantalum-aluminum alloy locally superheats the water-based ink beneath an exit nozzle to over 250°C. Within 5 μs, a bubble forms with peak pressures reaching 1.4 MPa (200 psi) and begins to expel ink out of the orifice. After 15 μs, the ink droplet, with a volume on the order of 10^{-10} liter, is ejected from the nozzle [37]. Within 24 μs of the firing pulse, the tail of the ink droplet separates, and the bubble collapses inside the nozzle, resulting in high cavitation pressure. Within less than 50 μs, the chamber refills, and the ink meniscus at the orifice settles.

The fabrication process of Hewlett-Packard inkjet heads has evolved as the printing resolution has increased. While the exact process flow is proprietary, a representative process follows. Fabrication starts with a silicon wafer, which is oxidized for thermal and electrical isolation (see Figure 4.31) [40]. Approximately 0.1 μm of tantalum-aluminum alloy is sputtered on, followed by aluminum containing a small amount of copper. The TaAl is resistive and has a near-zero thermal coefficient of expansion [38], while Al is a good conductor. The aluminum and TaAl are patterned, leaving an Al/TaAl sandwich to form conductive traces. Aluminum is then removed only from the resistor location to leave TaAl resistors. The resistors and conductive traces are protected by layers of PECVD silicon nitride, which is an electrical insulator, and PECVD silicon carbide, which is electrically conductive at elevated temperatures but is more chemically inert than the silicon nitride. The

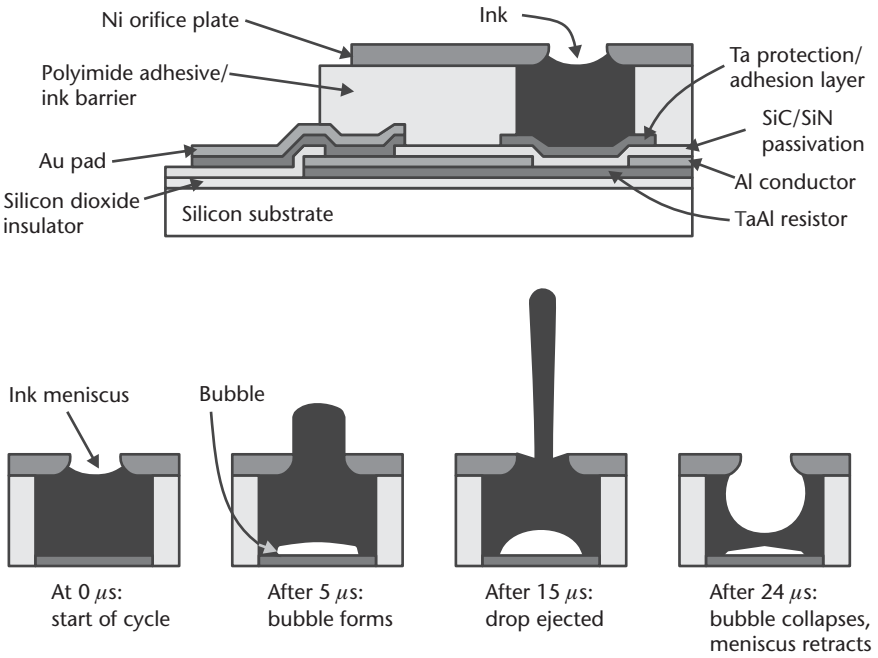


Figure 4.30 Concept of a Hewlett-Packard thermal inkjet head and the ink firing sequence. (*After:* [38, 39].)

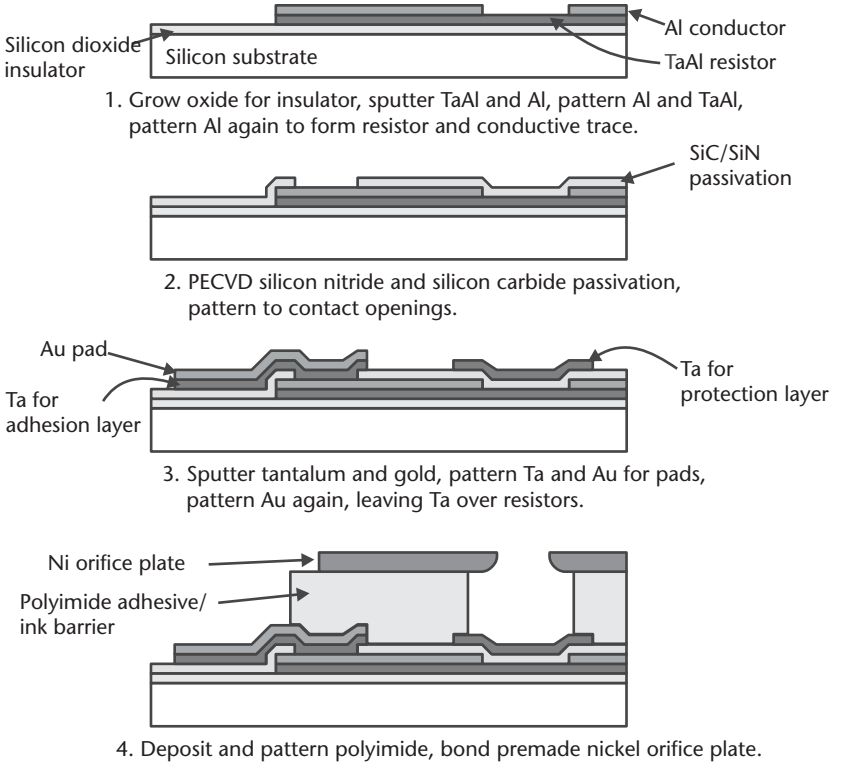


Figure 4.31 Fabrication process for Hewlett-Packard thermal inkjet head.

PECVD is performed at a sufficiently low temperature so as not to affect the metal already on the wafer. The use of this bilayer passivation, in addition to providing the appropriate thermal properties and needed chemical protection, reduces the incidence of pinholes. The SiC/SiN layers are patterned to make openings over the bond pads. Later generations of heads use an additional layer of tantalum, which is very hard, over the SiC/SiN in the resistor area to protect the underlying areas from the high cavitation pressure (up to 13 MPa) felt during bubble collapse, greatly lengthening the lifetime [40]. The tantalum sputtering is followed by gold sputtering without breaking vacuum. The Ta also acts as an adhesion layer for the Au. The Au and Ta are patterned, so they only remain on the contact pads and resistor. The gold is then etched off of the resistor, leaving it only on the bond pads. Next, polyimide is spun on, partially cured, and patterned to leave a channel through which ink flows to the resistor. The nickel orifice plate, which was separately fabricated using electroforming or laser ablation, is aligned and bonded to the silicon structure by the polyimide. Finally, the wafer is cut up to the final product size.

As the print resolution of HP thermal inkjet heads increased from 96 dpi in 1984 to 180 dpi to 300 dpi to 300 × 600 dpi in the mid 1990s, the number of heater resistors increased from 12 to 30 to 50 to 104, respectively [36, 37, 40]. For the earlier generations, there was one external contact pad to drive each resistor, plus several common grounds. For any electronic device, one of the greatest areas of reliability concern is where electrical contacts are made, and disposable inkjet heads are a particular concern because they are installed by the consumer. To improve reliability, as

well as reduce the area of silicon required for the output pads (which are large), on-chip *n*-channel metal-oxide-semiconductor (NMOS) driver circuitry was implemented on the 104-resistor chip, reducing the number of pads to 36. While the NMOS transistors require additional chip area and additional processing steps, the space savings is great enough that the overall chip cost is reduced [40].

Micromachined Valves

A new generation of miniature valves with electronic control would be desirable among both manufacturers and users of valves. For example, recent trends in home appliances indicate a shift towards total electronic control [41]. Electronically programmable gas stoves, currently under development, require low-cost, electronically controlled gas valves. Moreover, miniature valves are useful for the control of fluid flow functions in portable biochemical analysis systems [42] and automotive braking systems [43]. Some potential applications for silicon micromachined valves include:

- Electronic flow regulation of refrigerant for increased energy savings;
- Electronically programmable gas cooking stoves;
- Electronically programmable pressure regulators for gas cylinders;
- Accurate mass flow controllers for high-purity gas delivery systems;
- Accurate drug delivery systems;
- Control of fluid flow in portable biochemical analysis systems;
- Portable gas chromatography systems;
- Proportional control for electrohydraulic braking (EHB) systems.

The field of micromachined valves remains nascent. In order for silicon micromachined valves to gain a substantial foothold in the market, they must effectively compete with the relatively mature traditional valve technologies. These cover a broad range of actuation methods, media handling, pressures, flow rates, and price. It is unlikely that micromachined valves will displace traditional valves; rather, they will complement them in special applications where size and electronic control are beneficial.

The following sections describe three micromachined valves. The devices are from Redwood Microsystems, Inc., of Menlo Park, California; TiNi Alloy Company of San Leandro, California; and Alumina Micro, LLC of Bellingham, Washington. They illustrate the efforts of three small companies in commercializing microvalves. The first two valves operate on the principle of blocking a vertical fluid port with a silicon plug suspended from a spring that is sufficiently compliant to allow vertical displacement during actuation. Accordingly, the inlet pressure limit is relatively low, typically, less than 150 psig[2] (~ 1 MPa). The third valve uses an elegant pressure balancing scheme to reduce the pressure forces against an actuated silicon element thus increasing the inlet pressure to nearly 1,000 psig (~ 7 MPa).

2. The *psig* is a unit of differential (gauge) pressure equal to one psi (or 6.9 kPa).

Micromachined Valve from Redwood Microsystems

Early development of this valve took place in the mid 1980s at Stanford University [44]. Redwood Microsystems was founded shortly thereafter with the objective of commercializing the valve. The actuation mechanism of either normally open or normally closed valves[3] depends on the electrical heating of a control liquid sealed inside a cavity. When the temperature of the liquid rises, its pressure increases, thus exerting a force on a thin diaphragm wall and flexing it outward.

In a normally open valve, the diaphragm itself occludes a fluid port by its flexing action, hence blocking flow (see Figure 4.32). Upon removal of electrical power, the control liquid entrapped in the sealed cavity cools down, and the diaphragm returns to its flat position, consequently allowing flow through the port. The flexing membrane is in intimate contact with the fluid flow, which increases heat loss by conduction and severely restricts the operation of the valve. A more recent demonstration from Redwood Microsystems shows a thermal isolation scheme using a glass plate between the heated control liquid and the flexible membrane. Small perforations in the isolation glass permit the transmission of pressure to actuate the diaphragm.

The normally closed valve uses mechanical levering activated by a liquid-filled thermo pneumatic actuator to open an outlet orifice. The outward flexing action of the diaphragm under the effect of internal pressure develops a torque about a silicon fulcrum. Consequently, the upper portion of the valve containing the actuation element lifts the valve plug above the valve seat, permitting flow through the orifice.

The pressure that develops inside the sealed cavity results from the heating of the control liquid, which must meet several criteria in order to yield efficient actuation. In particular, the control liquid must be inert and noncorrosive. It must be electrically insulating but thermally conductive and must boil or expand considerably when heated. Redwood Microsystems uses one of the Fluorinert™ perfluorocarbon

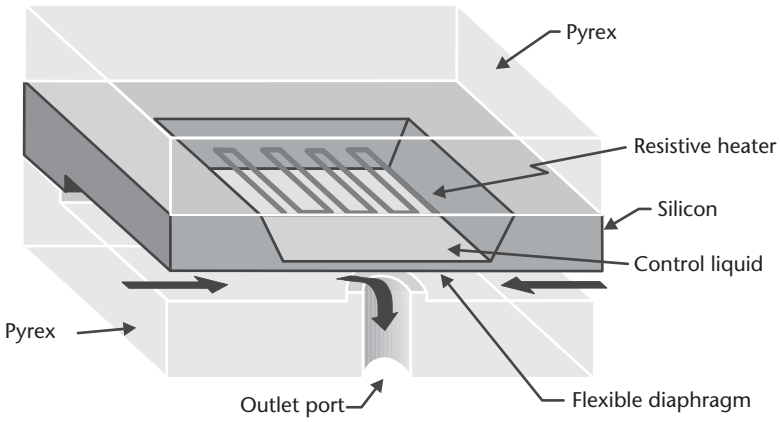


Figure 4.32 Illustration of a normally open valve from Redwood Microsystems. Heating of a control liquid sealed inside a cavity causes a thin silicon diaphragm to flex and block the flow through the outlet orifice. The inlet port is not shown.

3. The trademark name of the valve is the Fluistor™, short for fluid transistor because the valve is electrically gated in a fashion similar to the electronic transistor.

liquids from 3M Chemicals of St. Paul, Minnesota. Their boiling point ranges from 56° to 250°C, and they exhibit large temperature coefficients of expansion (~ 0.13% per degree Celsius). They are also electrically insulating and have a high dielectric constant. Clearly, the choice of control liquid determines the actuation temperature and, correspondingly, the power consumption and switching times of the valve.

The NO-1500 Fluistor normally open gas valve provides proportional control of the flow rate for noncorrosive gases. The flow rate ranges from 0.1 sccm up to 1,500 sccm. The maximum inlet supply pressure is 690 kPa (100 psig)[4], the switching time is typically 0.5s, and the corresponding average power consumption is 500 mW. The NC-1500 Fluistor is a normally closed gas valve (see Figure 4.33) with similar pressure and flow ratings, but its switching response is 1s and it consumes 1.5W. Because the Fluistor relies on the absolute temperature—rather than a differential temperature—of the control liquid for actuation, the valve cannot operate at elevated ambient temperatures. Consequently, the Fluistor is rated for operation from 0° to 55°C. The normally closed valve measures approximately 6 mm × 6 mm × 2 mm and is packaged inside a TO-8 can with two attached tubes (see Chapter 8). The packaging is further discussed in Chapter 8.

U.S. Patent 4,966,646 (October 30, 1990) describes the basic fabrication steps for a normally open valve; however, the fabrication details of a normally closed valve are not publicly available. The following process delineates the general steps to fabricate a normally closed valve. The features in the intermediate silicon layer are fabricated by etching both sides of the wafer in potassium hydroxide. The front-side etch forms the cavity that will later be filled with the actuation liquid. The etch on the bottom side forms the fulcrum as well as the valve plug. Accurate timing and a well-controlled etch rate of both etches ensure the formation of the thin

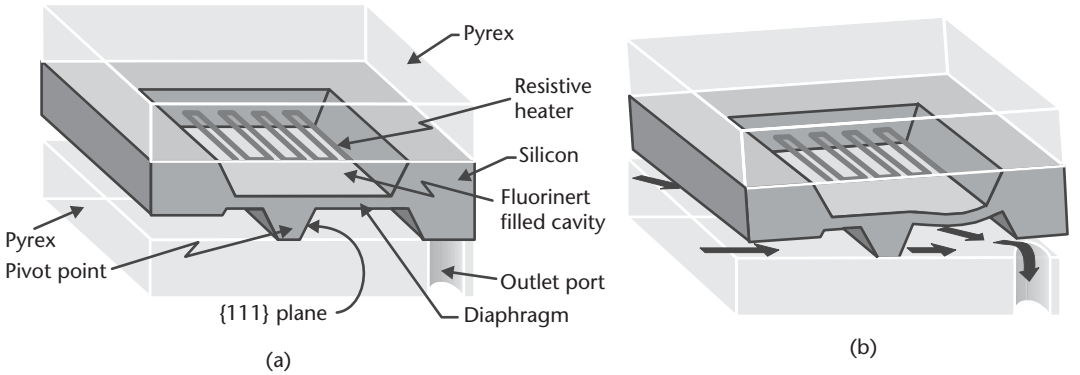


Figure 4.33 Illustration of the basic operating mechanism of a normally closed micromachined valve from Redwood Microsystems. (a) The upper stage of the valve normally blocks fluid flow through the outlet orifice. The inlet orifice is not shown. (b) Heating of the Fluorinert liquid sealed inside a cavity flexes a thin silicon diaphragm which in turn causes a mechanical lever to lift the valve plug. (*After:* Fluistor valve specification sheet of Redwood Microsytems of Menlo Park, California.)

4. Fluid flow through an ideal orifice depends on the differential pressure across it. The volume flow rate is equal to $C_D A_0 \sqrt{2\Delta P/\rho}$ where ΔP is the difference in pressure, ρ is the density of the fluid, A_0 is the orifice area, and C_D is the discharge coefficient, a parameter that is about 0.65 for a wide range of orifice geometries.

diaphragm in the middle of the silicon wafer. The top glass wafer is processed separately to form a sputtered thin-film metal heater. Ultrasonic drilling opens a fill hole through the top Pyrex glass substrate, as well as the inlet and outlet ports in the lower Pyrex glass substrate. Both glass substrates are sequentially bonded to the silicon wafer using anodic bonding. In the final step, the Fluorinert liquid fills the cavity. Special silicone compounds dispensed over the fill hole permanently seal the Fluorinert inside the cavity.

Micromachined Valve from TiNi Alloy Company

TiNi Alloy Company of San Leandro, California, is another small company with the objective of commercializing micromachined valves. Its design approach, however, is very different than that of Redwood Microsystems. The actuation mechanism relies on titanium-nickel (TiNi) [45], a shape-memory alloy—hence the name of the company. The rationale is that shape-memory alloys are very efficient actuators and can produce a large volumetric energy density, approximately five to 10 times higher than competing actuation methods. It is, however, the integration of TiNi processing with mainstream silicon manufacturing that remains an important hurdle.

The complete valve assembly consists of three silicon wafers and one beryllium-copper spring to maintain a closing force on the valve poppet (plug) (see Figure 4.34). One silicon wafer incorporates an orifice. A second wafer is simply a spacer defining the stroke of the poppet as it actuates. A third silicon wafer contains the valve poppet suspended from a spring structure made of a thin-film titanium-nickel alloy. A sapphire ball between a beryllium-copper spring and the third silicon wafer pushes the poppet out of the plane of the third wafer through the spacer of the second wafer to close the orifice in the first wafer. Current flow through the titanium-nickel alloy heats the spring above its transition temperature (~ 100°C),

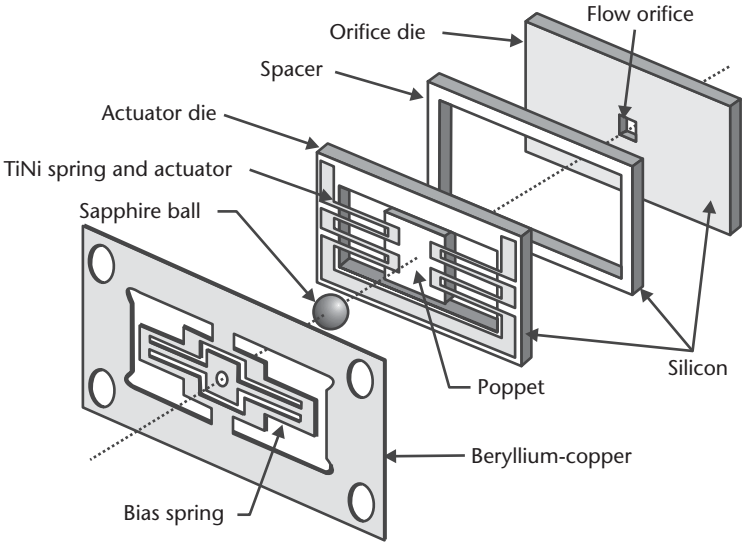


Figure 4.34 Assembly of the micromachined, normally closed valve from TiNi Alloy Company. The beryllium-copper spring pushes a sapphire ball against the silicon poppet to close the flow orifice. Resistive heating of the TiNi spring above its transition temperature causes it to recover its original flat (undeflected) shape. The actuation pulls the poppet away from the orifice, hence permitting fluid flow. (*After:* A. D. Johnson, TiNi Alloy Company of San Leandro, California.)

causing it to contract and recover its original undeflected position in the plane of the third wafer. This action pulls the poppet back from the orifice, hence permitting fluid flow.

The fabrication process relies on thin-film deposition and anisotropic etching to form the silicon elements of the valve (see Figure 4.35). The fabrication of the orifice and the spacer wafers is simple, involving one etch step for each. The third wafer containing the poppet and the titanium-nickel spring involves a few additional steps. Silicon dioxide is first deposited or grown on both sides of the wafer. The layer on the back side of the wafer is patterned. A timed anisotropic silicon etch using the silicon dioxide as a mask defines a silicon membrane. TMAH is a suitable etch solution because of its extreme selectivity to silicon dioxide. A titanium-nickel film, a few micrometers in thickness, is sputter deposited on the front side and subsequently patterned. Control of the composition of the film is critical, as this determines the transition temperature. Double-sided lithography is critical to ensure that the titanium-nickel pattern aligns properly with the cavities etched on the back side. Gold evaporation and patterning follows; gold defines the bond pads and the metal contacts to the titanium-alloy actuator. A wet or plasma etch step from the back side removes the thin silicon membrane and frees the poppet. At this point, the three silicon wafers are bonded together using a glass thermo-compression bond. Silicon fusion bonding is not practical because the titanium-nickel alloy rapidly oxidizes at temperatures above 300°C. Assembly of the valve elements remains manual, resulting in high production costs. The list price for one valve is about $200. Achieving wafer-level assembly is crucial to benefit from the cost advantages of volume manufacturing.

The performance advantage of shape-memory alloys manifests itself in low power consumption and fast switching speeds. The valve consumes less than 200 mW and switches on in about 10 ms and off in about 15 ms. The maximum gas flow rate and inlet pressure are 1,000 sccm and 690 kPa (100 psig), respectively. The valve measures 8 mm × 5 mm × 2 mm and is assembled inside a plastic package.

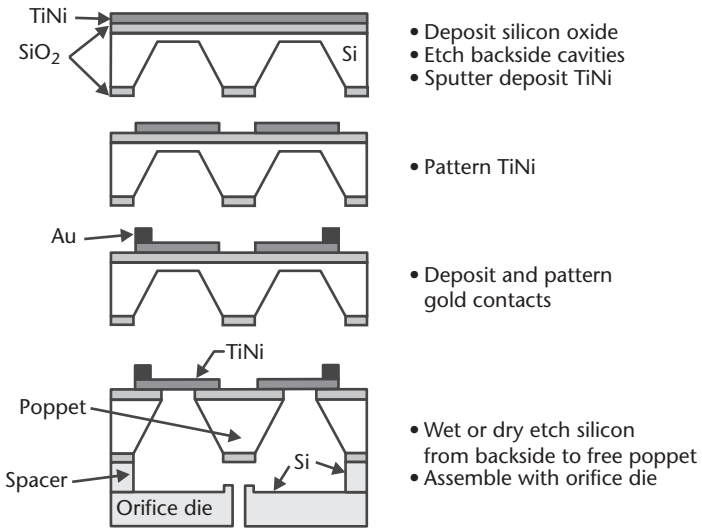


Figure 4.35 Fabrication sequence of the micromachined valve from TiNi Alloy Company. (*After:* [45].)

Sliding Plate Microvalve

Alumina Micro, LLC, of Bellingham, Washington, is developing micromachined valves under license based on technology developed at GE NovaSensor of Fremont, California. These valves are intended for use in such automotive applications as braking and air conditioning, which require the ability to control either liquids or gases at high pressures—as high as 2,000 psi (14 MPa)—over a wide temperature range (typically from –40°C to +125°C).

In micromachined valves that use a vertically movable diaphragm or plug over an orifice, such as the two examples discussed previously, the diaphragm or plug sustains a pressure difference across it. This pressure difference, when multiplied by the area, results in a force that must be overcome for the diaphragm to move. For high pressures and flow rates, the force becomes relatively large for a micromachined device. By contrast, the valve under development by Alumina Micro belongs to a family of valves known as sliding plate valves, in which a plate, or slider, moves horizontally across the vertical flow from an orifice. With appropriate design, the forces due to pressure can be balanced to minimize the force that must be supplied to the slider.

As shown in Figure 4.36, the valve is comprised of three layers of silicon [46, 47]. The inlet and outlets ports are formed in the top and bottom layers of silicon, respectively. For the normally open valve shown, fluid flows past the top controlling orifice formed between the slider and the top wafer, through the thickness of the second layer of silicon, and down out of the outlet port formed in the bottom wafer. Fluid flow also passes through the slot in the slider, under the slider, through the lower controlling orifice, and out of the outlet port. To reduce or turn off the flow, an actuator moves the slider to the right in the figure, reducing the area of the two controlling orifices. The pressure inside the slot is equal to the inlet pressure p_{in}. Therefore, the horizontal pressure forces acting on the internal surfaces of the slot are equal and opposite and balance each other. Similarly, the horizontal pressure forces acting on the external surfaces of the slot balance each other because the pressure outside the slot is equal to the outlet pressure p_{out}. The pressure forces are also balanced vertically, as the pressures on the top and bottom surfaces of the slider are equal to the inlet pressure [47]. In practice, small pressure imbalances due to flow are present, so some force is still required to move the slider, limiting operation to a few MPa (hundreds of psi).

The actuator is formed entirely in the middle silicon layer. There is a small (approximately 0.5 to 1 μm) gap above and below all moving parts to allow motion. The thermal actuator consists of a number of mechanically flexible "ribs" suspended in the middle and anchored at their edges to the surrounding silicon frame. Current flow through these electrically resistive ribs heats them, resulting in their expansion. The centers of the ribs push the movable pushrod to the left in the drawing [5], applying a torque about the fixed hinge and moving the slider tip in the opposite direction. When the current flow ceases, the ribs passively cool down by conduction of heat, both out the ends of the ribs and through the fluid. The mechanical restoring force of the hinges and ribs returns the slider to its initial position. Depending on the geometry of the actuator ribs, the actuation response time can vary from a few to hundreds of milliseconds. The depth of the recesses above and

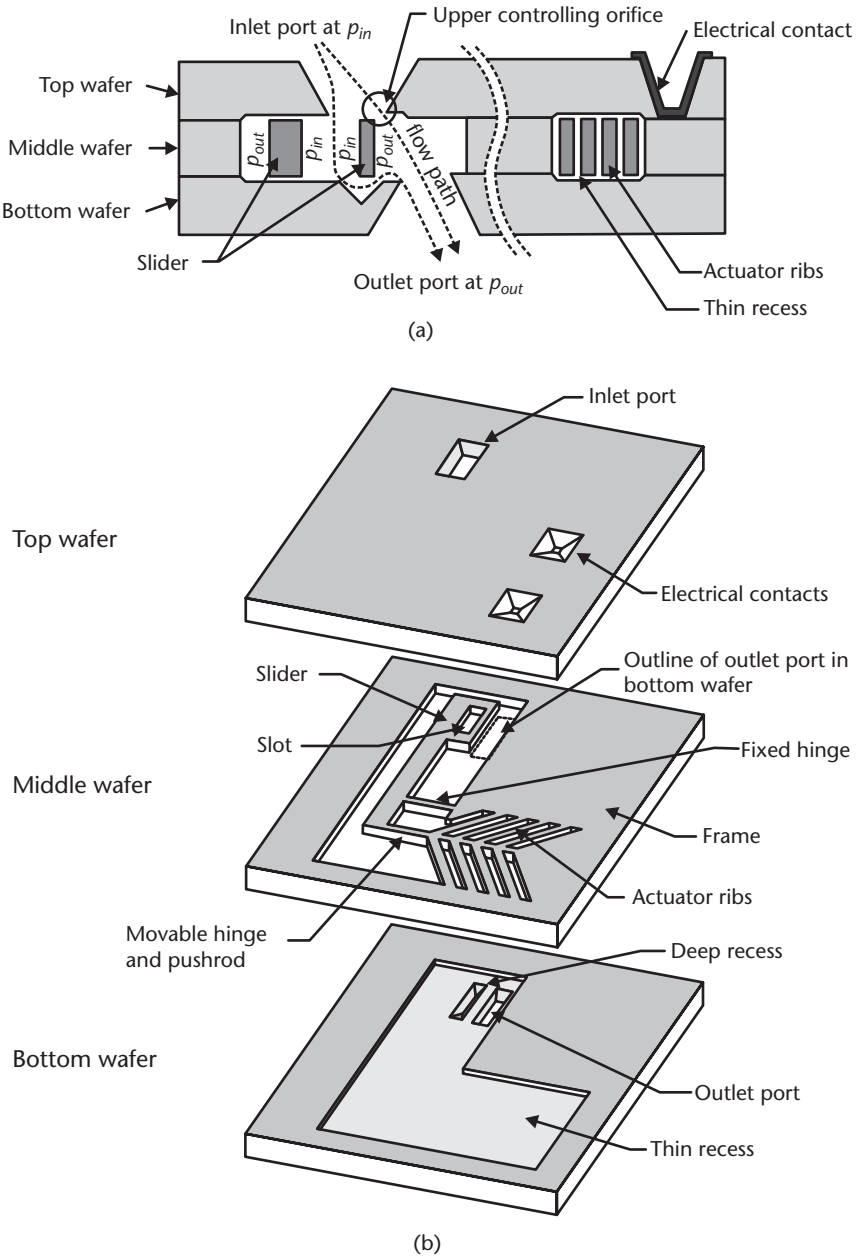


Figure 4.36 (a) A schematic cross section of the sliding plate microvalve depicting the inlet and outlet ports, as well as the slider and the ribs of the thermal actuator. The slider's motion to the right of the picture reduces the size of the upper and lower controlling orifices, therefore decreasing the flow through the valve. (b) A rendering of the three silicon wafers that comprise a micromachined pressure-balanced valve. The top and bottom wafers include the inlet and outlet ports, respectively. An intermediate wafer incorporates a thermal actuator that drives a slider suspended from two hinges.

below the ribs can be increased to lower the heat-flow rate. This reduces power consumption but slows the response when cooling.

This actuator design avoids rubbing parts, greatly improving the reliability of the valve. The force provided by the ribs can be raised by increasing both the silicon thickness and the number of ribs and can be on the order of one Newton, which is considered to be a very large force in micromachined structures. As the slider moves to the right, it reduces the areas of the upper and lower controlling orifices and thus the flow. Eventually, the slider closes off the orifices, and the flow drops to a negligible amount. A small amount of leakage occurs through the thin recess that is required to allow motion. In many applications, the leakage is considered small and is acceptable.

Because the ribs and the frame that constrains their ends are made of the same material (single-crystal silicon), the actuation force depends on the temperature gradient between them. Any changes in temperature that are uniform to the entire valve, such as fluctuations in the ambient temperature, cause both the ribs and the frame to expand and contract at the same rate, resulting in no actuation. This enables this valve design to operate over a very wide temperature range. The penalty for the use of an all-silicon valve is a much lower power efficiency, as silicon is a good thermal conductor (see Table 2.1) and heat is rapidly conducted out the ends of the ribs. A design advantage of using silicon is that the resistivity of the middle wafer can be specified by the designer over a range of several orders of magnitude, allowing the actuator resistance to be designed independently of the actuator dimensions.

To fabricate the valves, shallow recess cavities are etched in the top and bottom wafers for the clearances required for actuator motion. Etching in KOH creates the ports, deep recess, and through hole for electrical contacts (see Figure 4.36). These might also be formed using DRIE, but KOH etching is an inexpensive option and works well for this application. The actuator in the middle wafer is etched using DRIE. Aligned silicon fusion bonding combines the wafer stack. Metal is applied to the electrical contact areas of the middle wafer. Finally, the ports are protected with dicing tape to keep them clean, and the wafer is diced (described in Chapter 8). In use, the chips are held to the surface of a ceramic or metal package with an adhesive or solder and wire bonded.

A typical design may include ten or more rib pairs, where each pair is formed by two ribs connected in the middle to the pushrod. Each rib is approximately 100 μm wide, 2,000 μm long, and 400 μm thick, and is inclined at an angle of a few degrees. For an average temperature rise of 100°C, each rib pair contributes a force at the pushrod (and center of rib pair) of about 0.15N. The force falls nearly linearly to zero at the end of the stroke (about 5 to 10 μm). The lever structure formed by the fixed hinge and slider transform this large force and small displacement at the actuator to a moderate force and large displacement (>100 μm) at the tip of the slider near the fluid ports. The prototype valve initially demonstrated at GE NovaSensor [47] controlled water at pressures reaching 1.3 MPa (190 psig) and flows of 300 ml/min. Further design and fabrication improvements can increase these values to match the requirements of the automotive and industrial applications.

Micropumps

Micropumps are conspicuously missing from the limelight in the United States. By contrast, they receive much attention in Europe and Japan, where the bulk of the

development activities appears to be. An application for micropumps is likely to be in the automated handling of fluids for chemical analysis and drug delivery systems.

Stand-alone micropump units face significant competition from traditional solenoid or stepper-motor-actuated pumps. For instance, The Lee Company of Westbrook, Connecticut, manufactures a family of pumps measuring approximately 51 mm × 12.7 mm × 19 mm (2 in × 0.5 in × 0.75 in) and weighing, fully packaged, a mere 50g (1.8 oz). They can dispense up to 6 ml/min with a power consumption of 2W from a 12-V dc supply. But micromachined pumps can have a significant advantage if they can be readily integrated along with other fluid-handling components, such as valves, into one completely automated miniature system. The following demonstration from the Fraunhofer Institute for Solid State Technology of Munich, Germany [48], illustrates one successful effort at making a bidirectional micropump with reasonable flow rates.

The basic structure of the micropump is rather simple, consisting of a stack of four wafers (see Figure 4.37). The bottom two wafers define two check valves at the inlet and outlet. The top two wafers form the electrostatic actuation unit. The application of a voltage between the top two wafers actuates the pump diaphragm, thus expanding the volume of the pump inner chamber. This draws liquid through the inlet check valve to fill the additional chamber volume. When the applied ac voltage goes through its null point, the diaphragm relaxes and pushes the drawn liquid out through the outlet check valve. Each of the check valves comprises a flap that can move only in a single direction: The flap of the inlet check valve moves only as liquid enters to fill the pump inner chamber; the opposite is true for the outlet check valve.

The novelty of the design is in its ability to pump fluid either in a forward or reverse direction—hence its bidirectionality. At first glance, it appears that such a

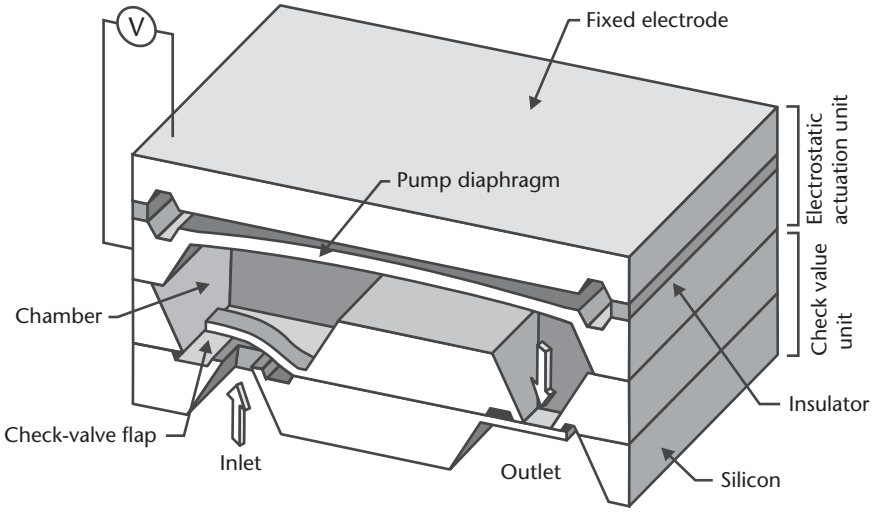


Figure 4.37 Illustration of a cutout of a silicon micropump from the Fraunhofer Institute for Solid State Technology of Munich, Germany [48]. The overall device measures 7 × 7 × 2 mm^3. The electrostatic actuation of a thin diaphragm modulates the volume inside a chamber. An increase in volume draws liquid through the inlet check valve. Relaxation of the diaphragm expels the liquid through the outlet check valve.

scenario is impossible because of the geometry of the two check valves. This is true as long as the pump diaphragm displaces liquid at a frequency lower than the natural frequencies of the two check valve flaps. But at higher actuation frequencies—above the natural frequencies of the flap—the response of the two flaps lags the actuation drive. In other words, when the pump diaphragm actuates to draw liquid into the chamber, the inlet valve flap cannot respond instantaneously to this action and remains closed for a moment longer. The outlet check valve is still open from the previous cycle and does not respond quickly to closing. In this instance, the outlet check valve is open and the inlet check valve is closed, which draws liquid into the chamber through the outlet rather than the inlet. Hence, the pump reverses its direction. Clearly, for this to happen, the response of the check valves must lag the actuation by at least half a cycle—the phase difference between the check valves and the actuation must exceed 180°. This occurs at frequencies above the natural frequency of the flap. If the drive frequency is further increased, then the displacement of theflaps becomes sufficiently small that the check valves do not respond to actuation.

The pump rate initially rises with frequency and reaches a peak flow rate of 800 μl/min at 1 kHz. As the frequency continues to increase, the time lag between the actuation and the check valve becomes noticeable. At exactly the natural frequency of the flaps (1.6 kHz), the pump rate precipitously drops to zero. At this frequency, the phase difference is precisely 180°, meaning that both check valves are simultaneously open—hence no flow. The pump then reverses direction with further increase in frequency, reaching a peak backwards flow rate of –200 μl/min at 2.5 kHz. At about 10 kHz, the actuation is much faster than the response of the check valves, and the flow rate is zero. For this particular device, the separation between the diaphragm and the fixed electrode is 5 μm, the peak actuation voltage is 200V, and the power dissipation is less than 1 mW. The peak hydrostatic back pressure developed by the pump at zero flow is 31 kPa (4.5 psi) in the forward direction and 7 kPa (1 psi) in the reverse direction.

The fabrication is rather complex, involving etching many cavities separately in each wafer and then bonding the individual substrates together to form the stack (see Figure 4.38). Etching using any of the alkali hydroxides is sufficient to define the cavities. The final bonding can be done by either gluing the different parts or using silicon fusion bonding.

Summary

This chapter presented a set of representative MEM structures and systems used in industrial and automotive applications, including a number of micromachined sensors, actuators, and a few passive devices. The basic sensing and actuation methods vary considerably from one design to the other, with significant consequences to the control electronics. Design considerations are many; they include the specifications of the end application, functionality, process feasibility, and economic justification.

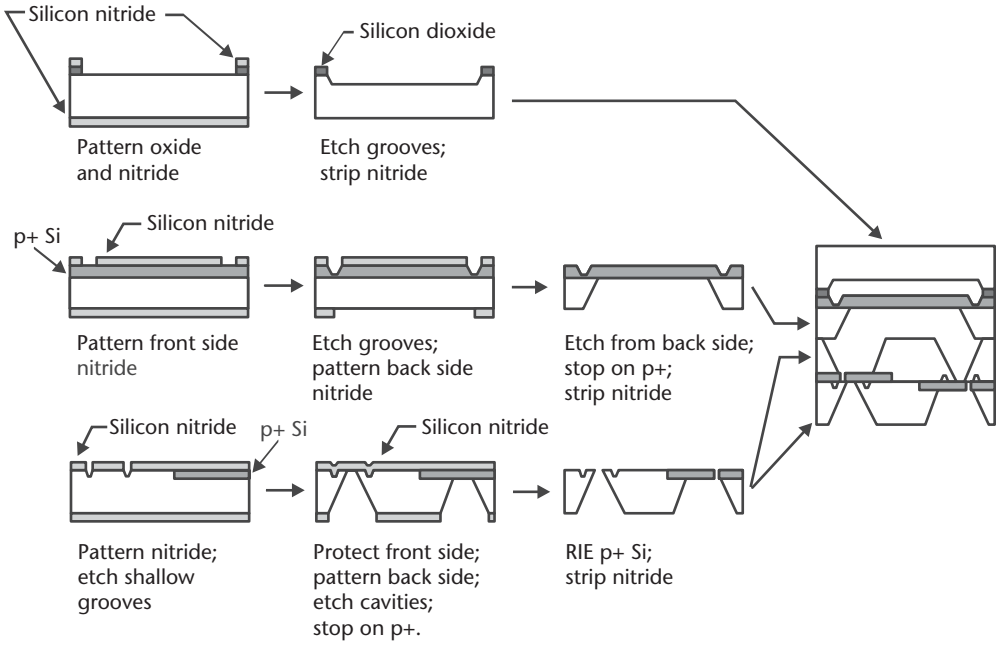


Figure 4.38 Fabrication process for an electrostatically actuated micropump.

References

[1] Yamazoe, N., and Y. Shimizu, "Humidity Sensors: Principles and Applications," *Sensors and Actuators*, Vol. 10, No. 3–4, November–December 1986, pp. 379–398.

[2] Kovacs, G. T. A., *Micromachined Transducers Sourcebook*, New York: McGraw-Hill, 1998, pp. 614–623.

[3] Johnson, W. A., and L. K. Warne, "Electrophysics of Micromechanical Comb Actuators," *Journal of Microelectromechanical Systems*, Vol. 4, No. 1, March 1995, pp. 49–59.

[4] Timoshenko, S., "Analysis of Bi-Metal Thermostats," *Journal of the Optical Society of America*, Vol. 11, 1925, p. 233.

[5] Klaassen, E. H., et al., "Silicon Fusion Bonding and Deep Reactive Ion Etching; A New Technology for Microstructures," *Proc. 8th Int. Conf. on Solid-State Sensors and Actuators*, Stockholm, Sweden, June 25–29, 1995, pp. 556–559.

[6] Halliday, D., and R. Resnick, *Physics*, Part Two, 3rd ed., Extended version, New York: Wiley, 1986, pp. 718–724.

[7] U.S. Patents 3,921,916, November 25, 1975, and 3,949,410, April 6, 1976.

[8] Peeters, E., and S. Verdonckt-Vandebroek, "Thermal Ink Jet Technology," *Circuits and Devices*, Vol. 13, No. 4, July 1997, pp. 19–23.

[9] Kneisel, L. L., J. D. Baker, and L. N. Goenka, "Silicon Micromachined CO_2 Cleaning Nozzle and Method," U.S. Patent 5,545,073, August 13, 1996.

[10] Siewell, G. L., W. R. Boucher, and P. H. McClelland, "The ThinkJet Orifice Plate: A Part with Many Functions," *Hewlett-Packard Journal*, Vol. 36, No. 5, May 1985, pp. 33–37.

[11] Wu, M. C., "Micromachining for Optical and Optoelectronic Systems," *Proceedings of the IEEE*, Vol. 85, No. 11, November 1997, pp. 1833–1856.

[12] Muller, R. S., and K. Y. Lau, "Surface-Micromachined Microoptical Elements and Systems," in *Integrated Sensors, Microactuators, & Microsystems (MEMS),* K. D. Wise (ed.), *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1705–1720.

[13] Pister, K. S. J., et al., "Microfabricated Hinges," *Sensors and Actuators*, Vol. A33, No. 3, June 1992, pp. 246–256.

[14] Czarnocki, W. S., and J. P. Schuster, "The Evolution of Automotive Pressure Sensors," *Sensors*, Vol. 16, No. 5, May 1999, pp. 52–65.

[15] NPC-107 data sheet, GE NovaSensor, 1055 Mission Court, Fremont, CA 94539, http://www.novasensor.com.

[16] Johnson, R. G., and R. E. Higashi, "A Highly Sensitive Silicon Chip Microtransducer for Air Flow and Differential Pressure Sensing Applications," *Sensors and Actuators*, Vol. 11, 1987, pp. 63–72.

[17] Roylance, L. M., and J. B. Angell, "A Batch Fabricated Silicon Accelerometer," *IEEE Trans. Electron Devices*, Vol. 26, No. 12, 1979, pp. 1911–1917.

[18] "Automotive Sensor Market," Market Report by Strategy Analytics, Inc., 199 Wells Avenue, Newton Centre, MA, 02459, February 2003.

[19] Suminto, J. T., "A Wide Frequency Range, Rugged Silicon Micro Accelerometer with Overrange Stops," *Proc. 9th Annual Int. Workshop on Micro Electro Mechanical Systems*, San Diego, CA, February 11–15, 1996, pp. 180–185.

[20] Sasayama, T., et al., "Highly Reliable Silicon Micro-Machined Physical Sensors in Mass Production," *Proc. 8th Int. Conf. on Solid-State Sensors and Actuators*, Stockholm, Sweden, June 25–29, 1995, pp. 687–690.

[21] Chau, K. H. –L., et al., "An Integrated Force-Balanced Capacitive Accelerometer for Low-G Applications," *Proc. 8th Int. Conf. on Solid-State Sensors and Actuators*, Stockholm, Sweden, June 25–29, 1995, pp. 593–596.

[22] Offenberg, M., et al., "Novel Process for a Monolithic Integrated Accelerometer," *Proc. 8th Int. Conf. on Solid-State Sensors and Actuators*, Stockholm, Sweden, June 25–29, 1995, pp. 589–592.

[23] Van Drieënhuizen, B. P., et al., "Force-Balanced Accelerometer with mG Resolution, Fabricated Using Silicon Fusion Bonding and Deep Reactive Ion Etching," *Proc. 1997 Int. Conf. on Solid-State Sensors and Actuators*, Chicago, IL, June 16–19, 1997, Vol. 2, pp. 1229–1230.

[24] "Inertial Technology for the Future," R. R. Ragan (ed.), *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-20, No. 4, July 1984, pp. 414–444.

[25] Emiliani, C., *The Scientific Companion*, 2nd ed., New York: Wiley, 1995, pp. 204–205.

[26] Beer, F. P., and E. R. Johnston, Jr., *Vector Mechanics for Engineers: Dynamics*, 3rd ed., New York: McGraw-Hill, 1977, pp. 716–719.

[27] Söderkvist, J., "Micromachined Gyroscopes," *Sensors and Actuators*, Vol. A43, 1994, pp. 65–71.

[28] Yazdi, N., F. Ayazi, and K. Najafi, "Micromachined Inertial Sensors," in *Integrated Sensors, Microactuators, & Microsystems (MEMS),* K. D. Wise (ed.), *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1640–1659.

[29] Chang, S., et al., "An Electroformed CMOS Integrated Angular Rate Sensor," *Sensors and Actuators*, Vol. A66, 1998, pp. 138–143.

[30] Langdon, R. M., "The Vibrating Cylinder Gyroscope," *The Marconi Review*, Fourth Quarter, 1982, pp. 231–249.

[31] Voss, R., et al., "Silicon Angular Rate Sensor for Automotive Applications with Piezoelectric Drive and Piezoresistive Read-Out," *Proc. 1997 Int. Conf. on Solid-State Sensors and Actuators*, Chicago, IL, June 16–19, 1997, Vol. 2, pp. 879–882.

[32] Lutz, M., et al., "A Precision Yaw Rate Sensor in Silicon Micromachining," *Proc. 1997 Int. Conf. on Solid-State Sensors and Actuators*, Chicago, IL, June 16–19, 1997, Vol. 2, pp. 847–850.

[33] Trautweiler, S., N. Mosier, and E. Zdankiewicz, "New Silicon-Based Metal-Oxide Chemical Sensors," *Sensors Magazine*, Vol. 16, No. 9, September 1999.

[34] Lyle, R. P., and D. Walters, "Commercialization of Silicon-Based Gas Sensors," *Proc. 1997 Int. Conf. on Solid-State Sensors and Actuators*, Chicago, IL, June 16–19, 1997, Vol. 2, pp. 975–978.

[35] Nguyen, N. -T., and S. T. Wereley, *Fundamentals and Applications of Microfluidics*, Norwood, MA: Artech House, 2002.

[36] Bhaskar, E. V., and J. S. Aden, "Development of the Thin-Film Structure for the ThinkJet Prinhead," *Hewlett-Packard Journal*, Vol. 36, No. 5, May 1985, pp. 27–33.

[37] Askeland, R. A., W. D. Childers, and W. R. Sperry, "The Second-Generation Thermal Ink-Jet Structure," *Hewlett-Packard Journal*, Vol. 39, No. 4, August 1988, pp. 28–31.

[38] Beatty, C. C., "A Chronology of Thermal Ink-Jet Structures," *Tech. Digest Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, June 3–6, 1996, pp. 200–204.

[39] Allen, R. R., J. D. Meyer, and W. R. Knight, "Thermodynamics and Hydrodynamics of Thermal Ink Jets," *Hewlett-Packard Journal*, Vol. 36, No. 5, May 1985, pp. 21–26.

[40] Aden, J. S., et al., "The Third Generation HP Thermal InkJet Printhead," *Hewlett-Packard Journal*, Vol. 45, No. 1, February 1994, pp. 41–45.

[41] "Under-Glass Controls: Creative Cooktop Delivers High Power and Fast Boiling," *Appliance Manufacturer*, July 1996, pp. 61–63.

[42] Anderson, R. C., et al., "Genetic Analysis Systems: Improvements and Methods," *Tech. Digest Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, June 8–11, 1998, pp. 7–10.

[43] U.S. Patent 6,533,366, March 18, 2003.

[44] U.S. Patents 4,824,073, April 25, 1989 and 4,966,646, October 30, 1990.

[45] Johnson, A. D., and E. J. Shahoian, "Recent Progress in Thin Film Shape Memory Microactuators," *Proc. IEEE Micro Electro Mechanical Systems*, Amsterdam, the Netherlands, January 29–February 2, 1995, pp. 216–219.

[46] U.S. Patent 6,523,560 B1, February 25, 2003, and U.S. Patent application publication US2002/0174891 A1, November 28, 2002.

[47] Williams, K. R., et al., "A Silicon Microvalve for the Proportional Control of Fluids," *Digest of the 10th International Conference on Solid-State Sensors and Actuators (Transducers '99)*, Sendai, Japan, June 1999, pp. 1804–1807.

[48] Zengerle, R., et al., "A Bidirectional Silicon Micropump," *Proc. IEEE Micro Electro Mechanical Systems*, Amsterdam, the Netherlands, January 29–February 2, 1995, pp. 19–24.

Selected Bibliography

Frank, R., *Understanding Smart Sensors*, Norwood, MA: Artech House, 1996.

Gad-El-Hak, M., *The MEMS Handbook*, Boca Raton, FL: CRC Press, 2001.

Kovacs, G. T. A., *Micromachined Transducers Sourcebook*, New York, NY: McGraw-Hill, 1998.

Micromechanics and MEMS: Classic and Seminal Papers to 1990, W. Trimmer (ed.), New York: IEEE, 1997.

Senturia, S. D., *Microsystem Design*, Boston, MA: Kluwer Academic Publishers, 2000.

Soloman, S., *Sensors Handbook*, New York: McGraw-Hill, 1998.

Wise, K. D. (ed.) "Special Issue on Integrated Sensors, Microactuators, and Microsystems (MEMS)," *Proceeding of the IEEE*, Vol. 86, No. 8, August 1998.

CHAPTER 5

MEM Structures and Systems in Photonic Applications

> "Our Technology and Engineering Emmy Awards have always honored the innovators that move our industry forward."
>
> —*Peter O. Price, President of the National Television Academy on the Emmy® Award to Texas Instrument's Digital Light Processing™ (DLP) Technology, October 2003.*

The penetration of MEMS technology in photonic applications is one that evokes in many minds stories of success. What made MEMS successful is that in many instances, it enabled new functionality by miniaturizing and arraying optical elements. Two notable markets and applications have benefited greatly from MEMS: displays and optical fiber communications. In displays, the Digital Light Processing (DLP™) technology from Texas Instruments of Dallas, Texas, has become a standard in small- and large-screen projection of digital images, with the Digital Mirror Device™ (DMD) at its core. In fiber-optical communications, there are a myriad of MEMS-based components in tunable lasers, optical switches, and optical attenuators, all key elements in transmitting data through optical fibers. But in hearing these success stories, one should not forget that the systems that are enabled by these MEMS-based components are very complex and encompass in their operation a multitude of technologies, with MEMS being just one of them. It is the convergence of all of these technologies that makes them collectively a success.

This chapter first describes in detail three commercially available products in imaging applications: an infrared image sensor and two image-projection devices. It then provides detailed insight into the operation of four types of products used in fiber-optical telecommunications: tunable lasers, wavelength lockers, optical switches, and variable optical attenuators. These components source and manipulate light as it travels within an optical fiber carrying information.

Imaging and Displays

Infrared Radiation Imager

Demonstrations of micromachined infrared bolometers and sensors have existed for many years, but the uncooled two-dimensional infrared imaging array from Honeywell, Inc., of Minneapolis, Minnesota [1], stands out in the crowd and competes effectively with traditional designs involving cooled cameras based on group II-VI compound semiconductors.

The basic approach of the Honeywell design, described in U.S. Patent 6,621,083 B2 (September 16, 2003), achieves high sensitivity to radiation by providing extreme thermal isolation for a temperature-sensitive resistive element. Incident infrared radiation heats a suspended sense resistor, producing a change in its resistance that is directly proportional to the radiation intensity (see Figure 5.1). The two-level structure, consisting of an upper silicon nitride plate suspended over a substrate, provides a high degree of thermal isolation corresponding to a thermal conductance of merely 10^{-8} W/K. This value approaches the theoretical lowest limit of 10^{-9} W/K due to radiative heat loss. The square silicon nitride plate is 50 μm on a side and 0.5 μm thick. The thin (50- to 100-nm) resistive element rests on the silicon nitride and has a large temperature coefficient of resistance in the range of –0.2 to –0.3% per degree Celsius. In order to capture most or all of the incident radiation, the fill factor—the area covered by the sensitive element as a fraction of the overall pixel area—must approach unity. The gap between the suspended plate and the substrate is approximately 1.8 μm. The silicon nitride plate and a thin reflecting metal directly underneath it form a quarter-wave resonant cavity to increase infrared absorption at wavelengths near 10 μm—corresponding to the peak radiation from a black body near 20°C. A two-dimensional array of these pixels images activity at or near room temperature and is useful for night vision.

The basic fabrication process relies on a surface micromachining approach, but unlike the polysilicon surface micromachining process, it incorporates an organic layer, such as polyimide, as the sacrificial material. The fabrication of the pixels occurs after the fabrication of standard CMOS electronic circuits on the silicon substrate. In a typical array size of 240 × 336 pixels, it is nearly impossible to obtain individual leads to each element. The integrated electronics provide multiplexing as well as scan and readout operations.

The CMOS electronic circuits are fabricated first. The last step in the CMOS process ensures that the surface is planar. One approach is by chemical-mechanical polishing (CMP) of a silicon dioxide passivation layer. The fabrication of the sense pixels begins with the deposition and patterning of the bottom metal films of the two-level structure. The composition of the metal does not appear to be critical. In the next step, the 1.8-μm thick sacrificial layer is deposited. The public literature

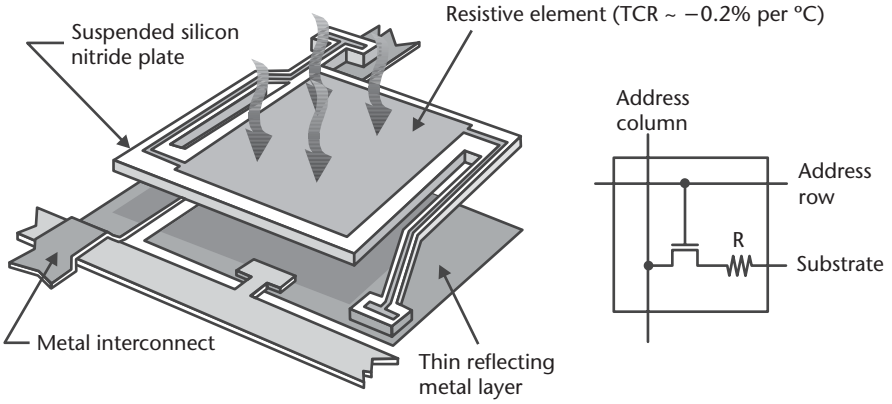


Figure 5.1 Illustration of a single sense element in the infrared imaging array from Honeywell. Incoming infrared radiation heats a sensitive resistive element suspended on a thin silicon nitride plate. Electronic circuits measure the change in resistance and infer the radiation intensity. (*After:* [1].)

does not specify the type of material, but one could use an organic polyimide film or photoresist that can sustain the subsequent thermal cycles of the fabrication process. Standard lithography and etching methods are applied to define contacts through the sacrificial layer to the underlying metal. These contacts also serve to form anchor points for the suspended plate. A 0.5-μm thick silicon nitride layer is deposited at low temperature and patterned using standard lithography in the shape of the suspended plate. The next deposition step is critical because it defines the thin temperature-sensitive resistor. Two families of materials exhibit suitable sensing properties:

- Vanadium oxides (VO_2, V_2O_3, and V_2O_5);
- Lanthanum manganese oxides ($La_{1-x}A_xMnO_3$; A = Ca, Sr, Ba, or Pb).

Sputtered vanadium oxides have a convenient sheet resistance (~25 kΩ per square at 25°C), acceptable 1/*f* noise, high absorption of infrared radiation, and a large TCR of about –0.2% per degree Celsius. Lanthanum manganese oxides yield even larger TCRs in the range of –0.3% per degree Celsius with low 1/*f* noise. The combination of low noise and high TCR is critical to increasing sensitivity. After the deposition and patterning of the resistive element, another silicon nitride layer is applied for encapsulation of the sensitive components. Removal of the sacrificial layer by plasma etching releases the silicon nitride plate. An oxygen plasma is effective at isotropically removing organic materials, including polyimide and photoresist. Finally, the parts are diced, then packaged under vacuum (<10 Pa residual pressure) to reduce heat loss by conduction.

The readout electronics (see Figure 5.1) activate a column of pixels by applying a voltage to their corresponding address column; they then measure the current from each transistor as a constant pulse voltage sequentially scans the address rows for the pixels in the column. The estimated change in temperature for an incident radiation power of 10^{-8} W is only 0.1°C. The corresponding resistance change is a measurable –10Ω for a 50-kΩ resistor. The thermal capacity of a pixel is 10^{-9} J/K, determined by the very small thermal mass of the suspended plate. Consequently, the thermal response time, defined by the ratio of thermal capacity to thermal conductance, is less than 10 ms, sufficiently fast for most imaging applications. The signal-to-noise ratio is limited by thermal noise and *1/f* noise to about 49 dB. Special circuits perform a calibration step that subtracts from the active image the signal of a blank scene. The latter signal incorporates the effects of nonuniform pixel resistance across the array. An intermittent shutter provides the blank scene signal, therefore allowing continuous calibration.

Projection Display with the Digital Micromirror Device™

The Digital Micromirror Device™ (DMD) is a trademark of Texas Instruments of Dallas, Texas, which developed and commercialized this new concept in projection display technology referred to as Digital Light Processing™ (DLP). U.S. Patent 4,615,595 (October 7, 1986) describes the early structure of the DMD. The technology has since undergone continuous evolution and improvements. Texas Instruments first introduced its new product family of DLP-based projection systems in 1996.

The DMD consists of a two-dimensional array of optical switching elements (pixels) on a silicon substrate (see Figure 5.2) [2]. Each pixel consists of a reflective micromirror supported from a central post. This post is mounted on a lower metal platform—the yoke—itself suspended by thin and compliant torsional hinges from two stationary posts anchored directly to the substrate. Two electrodes positioned underneath the yoke provide electrostatic actuation. A 24-V bias voltage between one of the electrodes and the yoke tilts the mirror towards that electrode. The nonlinear electrostatic and restoring mechanical forces make it impossible to accurately control the tilt angle. Instead, the yoke snaps into a fully deflected position, touching a landing site biased at the same potential to prevent electrical shorting. The angle of tilt is limited by geometry to ±10° (the direction of the sign is defined by the optics). The restoring torque of the hinges returns the micromirror to its initial state once the applied voltage is removed. CMOS static random-access memory (SRAM) cells fabricated underneath the micromirror array control the individual actuation states of each pixel and their duration. The *OFF* state of the memory cell tilts the mirror by –10°, whereas the *ON* state tilts it by +10°. In the *ON* state, off-axis illumination reflects from the micromirror into the pupil of the projection lens, causing this particular pixel to appear bright (see Figure 5.3). In the other two tilt states (0° and –10°), an aperture blocks the reflected light giving the pixel a dark appearance. This beam-steering approach provides high contrast between the bright and dark states. Each micromirror is 16 μm square and is made of aluminum for high reflectivity in the visible range. The pixels are normally arrayed in two dimensions on a pitch of 17 μm to form displays with standard resolutions from 800 × 600 pixels (SVGA) up to 1,280 × 1,024 pixels (SXGA). The fill factor, defined as the ratio of reflective area to total area, is approximately 90%, allowing a seamless (continuous) projected image

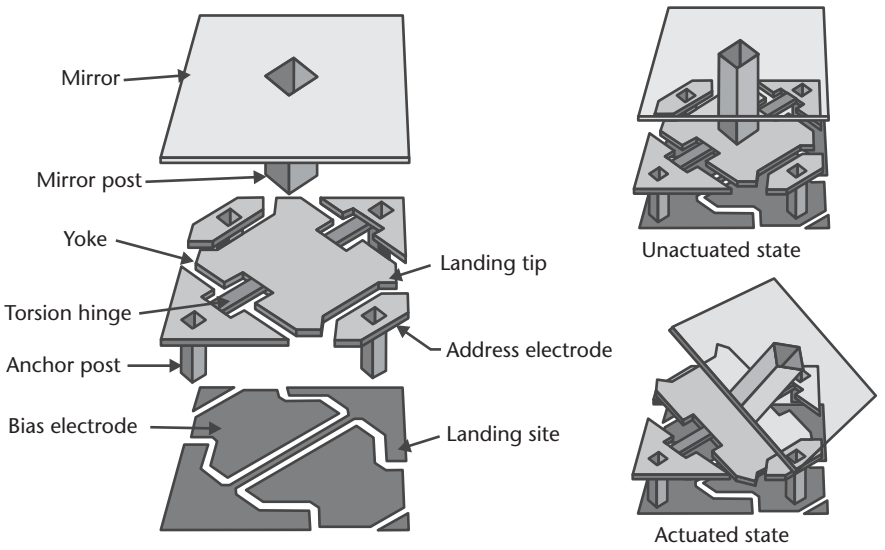


Figure 5.2 Illustration of a single DMD pixel in its resting and actuated states. The basic structure consists of a bottom aluminum layer containing electrodes, a middle aluminum layer containing a yoke suspended by two torsional hinges, and a top reflective aluminum mirror. An applied electrostatic voltage on a bias electrode deflects the yoke and the mirror towards that electrode. (*After:* [2].)

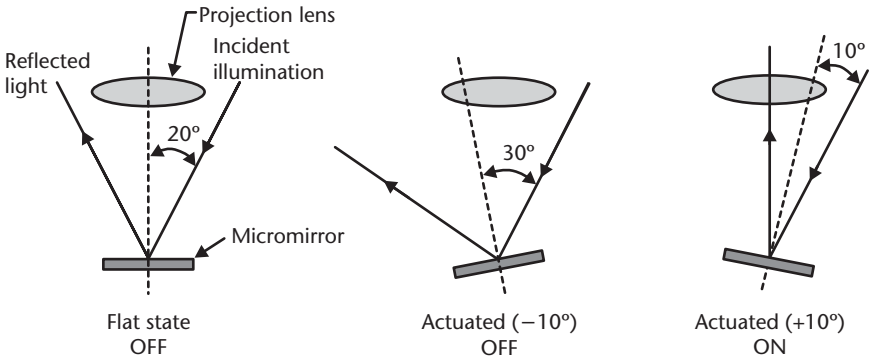


Figure 5.3 Illustration of optical beam steering using the switching of micromirrors. Off-axis illumination reflects into the pupil of the projection lens only when the micromirror is tilted in its +10° state, giving the pixel a bright appearance. In the other two states, the pixel appears dark [2].

free of pixelation. Texas Instruments reduced the pitch in 2000 from 17 μm to 14 μm in order to increase the number of available die per wafer and reduce cost.

While the operation of each mirror is only digital—in other words, the pixel is either bright or dark—the system is capable of achieving gray shades by adjusting the dwell time of each pixel—the duration it is bright or dark. The mechanical switching time including settling time is approximately 16 μs, much faster than the response of the human eye (on the order of 150 ms). At these speeds, the eye can only interpret the average amount—not the duration—of light it receives in a pulse. This, in effect, is equivalent to the impulse response of the eye. Modulating the duration of the pulse, or the dwell time, gives the eye the sensation of gray by varying the integrated intensity. Because the pixel switching speed is approximately 1,000 times faster than the eye's response time, it is theoretically possible to fit up to about 1,000 gray levels (equivalent to 10 bits of color depth). In actuality, full-color projection uses three DMD chips, one for each primary color (red, green, and blue), with each chip accommodating 8-bit color depth for a total of 16 million discrete colors. Alternatively, by using filters on a color wheel, the three primary colors can be switched and projected using a single DMD chip.

Texas Instruments uses surface micromachining to fabricate the DMD on wafers incorporating CMOS electronic address and control circuitry (see Figure 5.4). The basics of the fabrication process are in some respects similar to other surface micromachining processes; the etching of one or more sacrificial layers releases the mechanical structures. It differs in that it must address the reliable integration of close to one million micromechanical structures with CMOS electronics.

All micromachining steps occur at temperatures below 400°C, sufficiently low to ensure the integrity of the underlying electronic circuits. Standard 0.8-μm, double-metal level, CMOS technology is used to fabricate control circuits and SRAM memory cells. A thick silicon dioxide layer is deposited over the second CMOS metal layer. A CMP of this silicon dioxide layer provides a flat starting surface for the subsequent building of the DMD structures. A third aluminum metal layer is sputter deposited and patterned to provide bias and address electrodes, landing pads, and electrical interconnects to the underlying electronics. Photoresist is spin deposited, exposed, developed, and hardened with ultraviolet (UV) light to

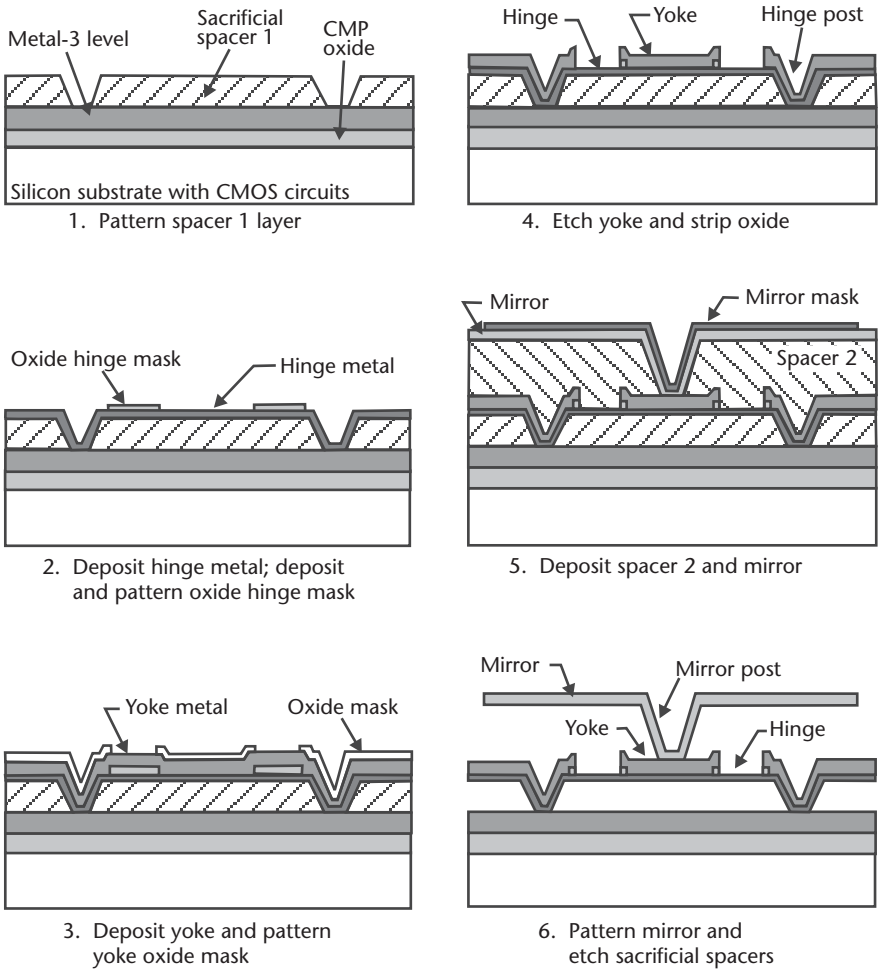


Figure 5.4 Fabrication steps of the Texas Instruments' DMD [2].

form the first sacrificial layer (sacrificial spacer 1). A sputter deposition of an aluminum alloy (98.8% Al, 1% Si, 0.2% Ti) defines the hinge metal layer. The mechanical integrity of the DMD relies on low stresses in the hinge. A thin silicon dioxide mask is then deposited with PECVD and patterned to protect the torsion hinge regions. The aluminum is not etched after this step. Retaining this silicon dioxide mask, another sputtering step deposits a thicker yoke metal layer, also made of a proprietary aluminum alloy. A thin layer of silicon dioxide is subsequently deposited and patterned in the shape of the yoke and anchor posts. An etch step removes the exposed aluminum areas down to the organic sacrificial layer. But in the regions where the oxide hinge mask remains, only the thick yoke metal is removed, stopping on the silicon dioxide mask and leaving intact the thin torsional hinges. Both silicon dioxide masking layers are stripped before a second sacrificial layer, also made of UV-hardened photoresist, is deposited and patterned. Yet another aluminum alloy sputter deposition defines the mirror material and the mirror post. A silicon dioxide mask protects the mirror regions during etch of the aluminum alloy.

The remaining fabrication steps address the preparation for sawing and packaging, made difficult by the delicate micromechanical structures. A wafer saw cuts the

silicon along edge scribe lines to a depth that allows breaking the individual dice apart at a later stage. An oxygen-plasma etch step removes both sacrificial layers and releases the micromirrors. A special passivation step deposits a thin, antistiction layer to prevent any adhesion between the yoke and the landing pads. Finally, a singulation process breaks apart and separates the individual dice. The packaging of the DMD is discussed in Chapter 8.

Reliability is the *sine qua non* of the commercial success of DMD technology. The designs described here are the result of extensive efforts at Texas Instruments aimed at understanding the long-term operation of the pixels as well as their failure modes. The DMD micromirrors are sufficiently robust to withstand normal environmental and handling conditions, including 1,500G mechanical shocks, because the weight of the micromirrors is insignificant. The major failure and malfunction mechanisms are surface contamination and hinge memory. The latter is the result of metal creep in the hinge material and causes the mirror to exhibit a residual tilt in the absence of actuation voltages. The reliability of the DMD is further discussed in Chapter 8.

Grating Light Valve™ Display

The Grating Light Valve™ (GLV) is a novel display concept invented initially at Stanford University. Silicon Light Machines of Sunnyvale, California, a division of Cypress Semiconductor Corp. of San Jose, California, is developing a commercial product based on the licensed technology [3]. The fundamental light-switching concept relies on closely spaced parallel rows of reflective ribbons suspended over a substrate (see Figure 5.5). The separation gap between the ribbons and the substrate is

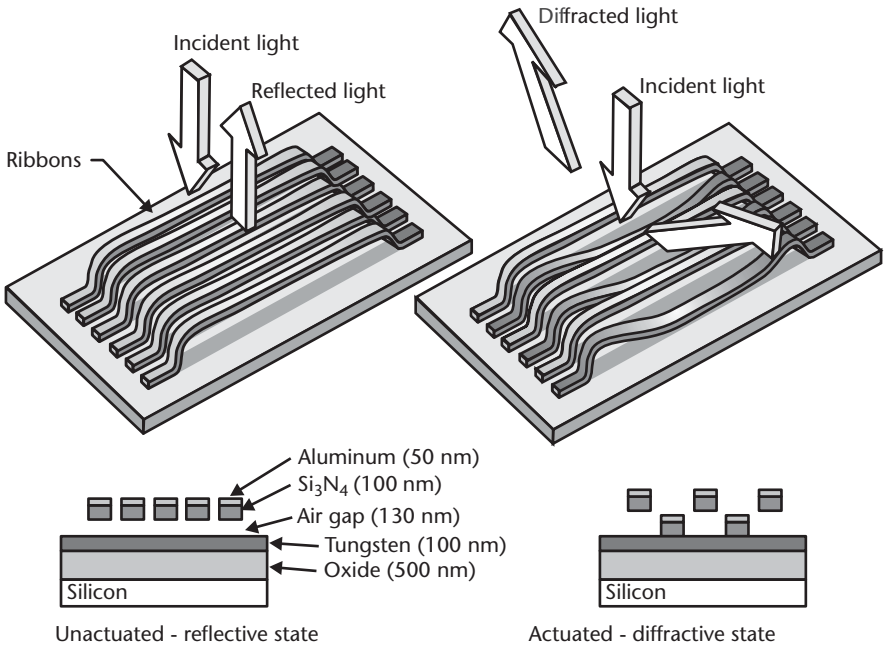


Figure 5.5 Illustration of the operating principle of a single pixel in the GLV. Electrostatic pull down of alternate ribbons changes the optical properties of the surface from reflective to diffractive. (*After:* [3].)

approximately one quarter the wavelength of light in the visible spectrum. In their resting state, the ribbons appear as a continuous surface to incident light, and normal reflection occurs. But when an electrostatic voltage pulls down alternate rows of ribbons, the light reflecting from the deflected ribbons travels an additional one half of a wavelength (twice the gap) and thus becomes 180° out of phase with respect to the light from the stationary ribbons. This effectively turns the ribbons into a phase grating, diffracting the incident light into higher orders. The angle of diffraction depends on the wavelength and the pitch—or periodicity—of the ribbons.

The entire display element consists of a two-dimensional array of square pixels, each approximately 20 μm on a side containing two fixed and two flexible ribbons. The mechanical structure of the ribbon relies on a thin silicon nitride film under tension to provide the restoring force in the absence of actuation. The reflecting surface is a 50-nm-thick aluminum layer. The underlying electrode is made of tungsten isolated from the substrate by silicon dioxide.

The optical projection system includes an aperture mounted over the display element (see Figure 5.6). Light-absorbing material surrounding the aperture blocks the reflected light but allows the first diffraction orders to be imaged by the projection lens. The incident illumination may be normal to the chip, sending the diffracted orders off axis. Alternatively, the use of off-axis illumination simplifies the imaging optics in a scheme similar to projection with the DMD described in the previous chapter.

For full color display, each pixel consists of three sets of ribbons, one for each of the three primary colors (red, green, and blue). The design of the pitch is such that the projection lens images the diffraction order of only one single color from each subpixel. The pitch of the red subpixel must be larger than that of green, and in turn larger than that of blue.

The GLV display supports at least 256 gray shades or 8-bit color depth by rapidly modulating the duration ratio of bright to dark states. This in turn varies the light intensity available for viewing—similar to the scheme used in the DLP by Texas Instruments. Early display prototypes demonstrated a contrast ratio between the

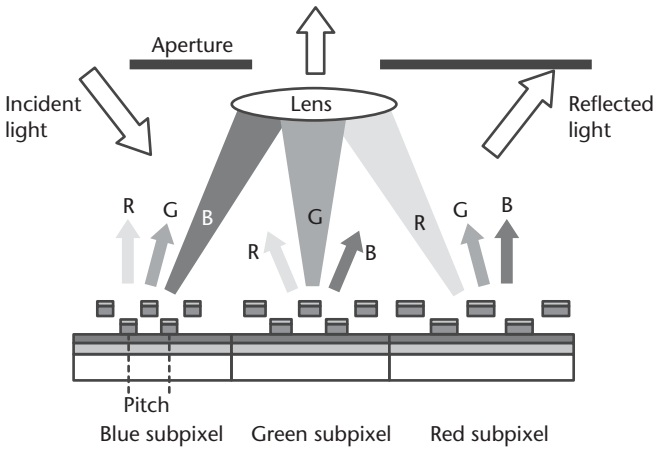


Figure 5.6 Implementation of color in a GLV pixel. The pitch of each color subpixel is tailored to steer the corresponding light to the projection lens. The aperture blocks the reflected light but allows the first diffraction order to enter the imaging optics. The size of the pixel is exaggerated for illustration purposes.

bright and dark states in excess of 200. The fill ratio—the percentage area available to reflect light—is approximately 70%, with a potential for further improvement by reducing the unused space between ribbons—the pitch, and not the spacing, determines the diffraction angle.

A key advantage of the GLV over other display technologies is its fast speed. The small size and weight of the ribbon, combined with the short stroke, provide a switching speed of about 20 ns, about one thousand times faster than the DMD. At these speeds, the address and support electronics become simple. There is no longer a need for fast memory buffers, such as those required for conventional active matrix liquid crystal displays, to compensate for the mismatch in speeds between the electronics and the display elements. Moreover, there is little power required to actuate the very small ribbons.

The very fast switching has also allowed Silicon Light Machines to explore a simpler scheme, whereby the projected image of a single row of pixels is rapidly scanned through the optics to build a two-dimensional picture. Projection at video rate for a high-resolution display requiring 1,000 horizontal lines implies a data scan rate of 60,000 lines per second. Incorporating 256 shades of gray increases the bit refresh rate to 15.4 MHz, which corresponds to a pixel switching every 65 ns—well within the capability of the GLV. This new scheme allows simplifying the GLV to a single row of pixels instead of a two-dimensional array and hence reduces associated manufacturing costs.

The fabrication involves the surface micromachining of the ribbons and their release by etching a sacrificial layer. The process begins with the deposition of an insulating 500-nm thick silicon dioxide layer over a silicon wafer, followed by the sputter deposition or CVD of tungsten. The tungsten is patterned using standard lithography and etched in SF_6-based plasma to define the electrodes for electrostatic actuation. The sacrificial layer is then deposited. The details of this layer are not publicly available, but many possibilities exist, including organic polymers. This layer is very thin, measuring approximately 130 nm, one quarter the wavelength of green light. Silicon nitride and aluminum are deposited next, followed by patterning in the shape of narrow ribbons. The release step is last. Oxygen plasma is useful for the removal of organic sacrificial layers, such as photoresist. It is also possible to consider using sputtered amorphous silicon as a sacrificial layer. Its selective removal, however, may require an exotic etch step involving xenon difluoride (XeF_2). This etchant sublimes at room temperature from its solid form and reacts spontaneously with silicon to form volatile SiF_4. Its advantage over SF_6 or CF_4 is that it does not require a plasma, and it does not etch silicon nitride, silicon oxide, or aluminum. But xenon difluoride is a hazardous chemical, reacting with water moisture to form hydrofluoric acid. It is not used in the integrated circuit industry.

Fiber-Optic Communication Devices

The rise and fall of scores of start-up companies during the bubble years (1997–2001) of the fiber-optic telecommunication industry left a legacy of technical innovations and novel designs, especially as relating to MEMS. In the economic downturn, many companies closed their doors, and it could be years before their intellectual property is applied to other fields. A few companies have survived and

continue to seek customers for their products. The difficult economic environment has necessitated that the surviving companies develop products that are cost competitive, especially against similar products made using alternative traditional technologies, while passing the stringent Telcordia™ standards of reliability (see Chapter 8). MEMS has become widely accepted as the fabrication technology of choice for a number of functions, in particular for dynamic attenuation of the light intensity inside the fiber, known as variable optical attenuators (VOAs); beam steering of light among an array of fibers, also known as optical switching or cross connects; and, to a lesser extent, as components within tunable lasers. It is important to note that while the primary market that drove the development of such devices was fiber-optic telecommunication, there remain other applications, albeit in smaller markets, that can benefit from these innovations (e.g., imaging, microscopy, and spectroscopy).

We examine in this section four different types of MEMS-based photonic devices whose sole function is to manipulate or generate light. We begin first with two distinct embodiments of a tunable laser product, one from Iolon, Inc., of San Jose, California, and the other from Santur Corporation of Fremont, California. Next, we describe a wavelength locker from Digital Optics Corporation of Charlotte, North Carolina. We then follow with an optical switch from Sercalo Microtechnology, Ltd., of Liechtenstein; then a beam steering mirror, or three-dimensional (3-D) optical switch, from Integrated Micromachines, Inc., of Irwindale, California; and finally a VOA from Lightconnect, Inc., of Newark, California.

Tunable Lasers

Lasers are at the core of fiber-optical communication where information is impressed upon streams of light inside a fiber. The advent of wavelength-division multiplexing (WDM) in the last decade offered a tremendous increase in information bandwidth by multiplexing multiple wavelengths into a single fiber. But as the number of wavelength channels increased to 100 and beyond, wavelength agility and the ability to switch between channels without human intervention has become of great importance. This is where tunable lasers promise to play a significant role [4].

Tunable lasers as bench-top test instruments have achieved a great degree of technical maturity in the recent past. Companies such as New Focus, Inc., of San Jose, California, and Agilent Technologies of Palo Alto, California, have offered such products for many years. But the innovation brought forth by MEMS technology aims to miniaturize these instruments from bench-top dimensions to fit in the palm of a hand. This miniaturization is necessary because equipment space inside central offices is very limited—a central office houses racks of electronic and optical equipment for processing and routing of data and voice. The use of tunable lasers in telecommunications has been primarily in the wavelength range of 1,528 nm to 1,565 nm (known as the C-Band) and 1,570 nm to 1,610 nm (known as the L-Band). The International Telecommunication Union (ITU) of Geneva, Switzerland, has specified the use to be on a grid of discrete channels throughout the C- and L-Bands at optical frequencies spaced 50 GHz (~ 0.4 nm) apart [5]. This grid specification brings forth the need for a wavelength "locker" to prevent a laser from drifting from its assigned wavelength on the ITU grid.

A basic laser consists of an optical amplification medium (a gain medium) positioned inside a resonant cavity [6, 7]. The amplification medium can be a gas (e.g., helium-neon or argon), a crystal (e.g., a ruby or neodymium), or, most commonly, a semiconductor material (e.g., GaAs, AlGaAs, or InP, depending on the wavelength of interest). The resonant cavity, in its simplest form, consists of two partially reflecting surfaces with an optical separation equal to an integral number of half wavelengths. Its role is to provide positive optical feedback by circulating light within its geometrical boundaries (see Figure 5.7).

In an optical amplifier, an electrical current or a high-intensity light excites (pumps) electrons from a low-energy (ground) state to a high-energy (excited) state. When the population of electrons in the excited state exceeds that in the ground state, the material reaches *population inversion* and becomes capable of a physical process known as stimulated amplification [8]. In this process, an incoming photon whose energy is equal to the energy difference between the excited and ground states *stimulates* the relaxation of an electron to its ground state, thus releasing a photon that is coherent (i.e., preserving the phase) and chromatic (i.e., preserving the wavelength) with the incident photon. When an optical amplifier is placed within an optically resonant cavity, light reflects back and forth inside the resonator with coherent amplification at every pass within the gain medium—it is this positive feedback that gives rise to the high intensity of the laser beam. However, resonance occurs only at certain specific wavelengths or frequencies—these are called the cavity longitudinal modes and are separated by a frequency equal to $c/2L$, where c is the speed of light within the medium and L is the optical cavity length [9]. At these frequencies, the optical length of the cavity is an integral number of half wavelengths. Light at other wavelengths rapidly decays. For relatively long cavities (>0.5 mm), multiple discrete modes coexist within the available spectrum of the gain medium, and the light

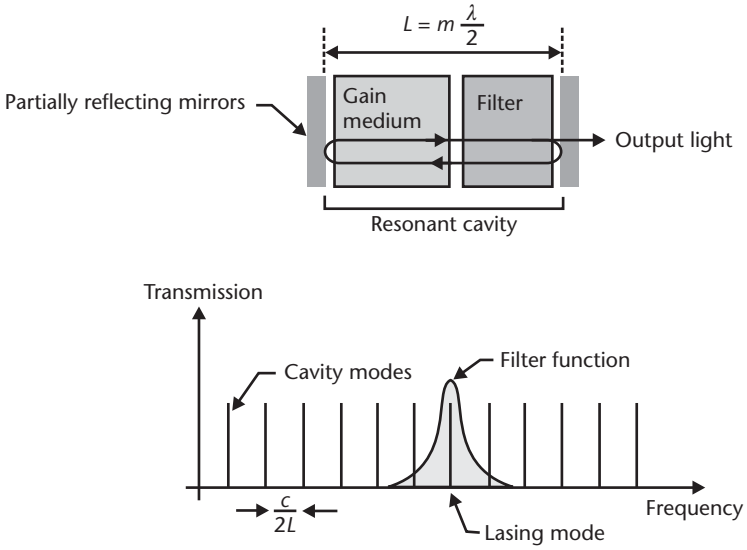


Figure 5.7 Illustration of the building blocks of a laser. A gain medium amplifies light as it oscillates inside a resonant cavity. Only select wavelengths called longitudinal cavity modes that are separated by a frequency equal to $c/2L$ may exist within the cavity. A wavelength filter with a narrow transmission function selects one lasing mode and ensures that the output light is monochromatic.

beam is not necessarily monochromatic, as multiple modes may participate in lasing. A wavelength filter, typically a grating, selects only one desired wavelength to generate a monochromatic laser beam (see Figure 5.7).

The tuning of a laser requires two simultaneous operations: the tuning of the filter to the new desired wavelength and the tuning of the optical length of the cavity such that one of the resonant longitudinal modes defined by $c/2L$ overlaps the desired wavelength (see Figure 5.7). Often referred to as the phase tuning, this is a condition for resonance. Additionally, at any output wavelength of a tunable laser, the amplification medium must possess a reasonable gain before lasing can occur—this is strictly a material property that dictates the choice of the material. The two lasers described here achieve the same objective using two radically different approaches. The Iolon approach achieves both tuning steps by using a MEMS-type microactuator [10]. The Santur approach [11] does it by heating and cooling the gain medium to change the index of refraction.

The main specifications of a tunable laser are wavelength in nanometers (or the corresponding optical frequency in Hz), tuning range in nanometers, spectral linewidth at the lasing frequency in Hz (the narrower the linewidth, the higher the coherence of the output beam), output optical power expressed in milliwatts or in dBm (the reference 0 dBm level is at 1 mW), relative intensity noise over a given frequency bandwidth (RIN) expressed in db/Hz, and side-mode suppression ratio (SMSR) in dB, which measures the power ratio at the lasing fundamental mode or wavelength to its nearest allowed mode. For applications in telecommunications, the specifications vary between short-distance (a few kilometers) and long-distance (>800 km) transmission. The latter requires more stringent specifications; for instance, the power is typically 13 ± 0.25 dBm (20 ± 1 mW) over the entire C-Band, the RIN needs to be lower than –120 dB/Hz, and the SMSR is higher than 45 dB.

The External Cavity Tunable Laser from Iolon

The laser design used by Iolon [10] belongs to a family of external-cavity lasers known after their inventors as Littman-Metcalf (see Figure 5.8) [12]. The three key building blocks are physically separate and hence can be optimized individually. External cavity lasers can also deliver superior properties in the form of stable power as well as high monochromaticity (measured as narrow line width) [13].

In this laser, the amplification medium consists typically of an InGaAsP/InP semiconductor diode with multiple quantum wells (a laser diode) because its gain spectrum covers the entire C-Band [14]. A thermoelectric cooler (TEC) maintains the temperature of the laser diode at approximately 25°C to increase diode lifetime and minimize chromatic thermal drift—the gain spectrum is a strong function of temperature. The wavelength filter is a glancing-angle ruled blazed or holographic grating [15] with a typical periodicity of 1,200 lines per millimeter. A partially reflective coating on one facet of the laser diode and a reflective mirror bound the external cavity [see Figure 5.8(a)]. With an effective cavity length of 8 mm, the spacing between the cavity modes is approximately 18 GHz (~ 0.2 nm) (i.e., nearly 190 distinct modes fit within the C-Band). The other facet of the laser diode must be highly transmissive (coated with an antireflective multilayer coating) in order to avoid forming a spurious resonant cavity within the diode itself—the reflectance is often significantly less than 10^{-3}. Light emanates from the laser diode through a

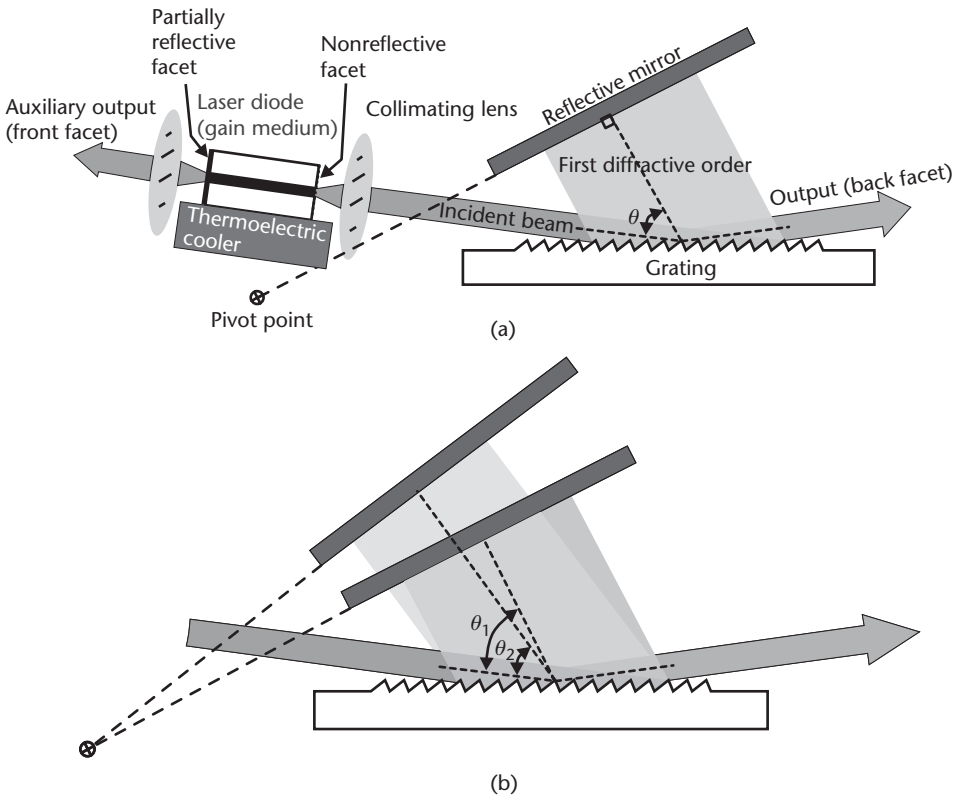


Figure 5.8 (a) Illustration of the Littman-Metcalf external cavity laser configuration. Light from the laser diode is collimated and diffracted by a grating acting as a wavelength filter. Lasing occurs only at one wavelength, whose diffraction order is reflected by the mirror back into the cavity. (b) Rotating the mirror around a virtual pivot point changes the wavelength and tunes the laser.

collimating lens, then diffracts on the grating. The mirror reflects back into the cavity and the gain medium only one wavelength whose diffracted beam is exactly perpendicular to the mirror. This wavelength (and corresponding diffraction angle θ) depends strictly on the grating pitch as well as the relative angle of the mirror with the diffraction grating. In actuality, because the diffraction grating has finite dispersion [16], the linewidth of the reflected wavelength is broadened to a few picometers. The output of the laser is typically the main (undiffracted) order reflecting from the grating, but an auxiliary output can be taken from the partially reflective facet of the laser diode.

The Littman-Metcalf configuration utilizes a fixed grating but rotates the reflective mirror to tune the laser to a different wavelength [Figure 5.8(b)]. It is the rotation of the mirror that achieves both tuning operations simultaneously: it selects a different diffracted wavelength from the grating, and it modulates the physical length of the cavity. By appropriately selecting a virtual pivot point [17], the dimensional change of the cavity length can be such that an integral number of *new* half wavelengths can fit within the cavity—note that a rotation about a virtual pivot point is geometrically equivalent to a rotation about a real pivot point and a linear translation. A poor choice of pivot point or misalignment can cause serious

degradation in the performance and tunability of the laser. One such degradation is mode hopping, when the cavity no longer supports an integral number of half wavelengths, causing the laser to "hop" to a different cavity mode (or wavelength) that satisfies the resonance condition.

Past designs of Littman-Metcalf-type lasers incorporated large traditional actuators, such as piezoelectric rods or voice-coil actuators, to rotate the mirror around the virtual pivot point. The Iolon approach uniquely incorporates an electrostatic rotary microactuator to miniaturize the overall size of the laser (see Figure 5.9). A gold-coated silicon substrate mounted vertically on top of special

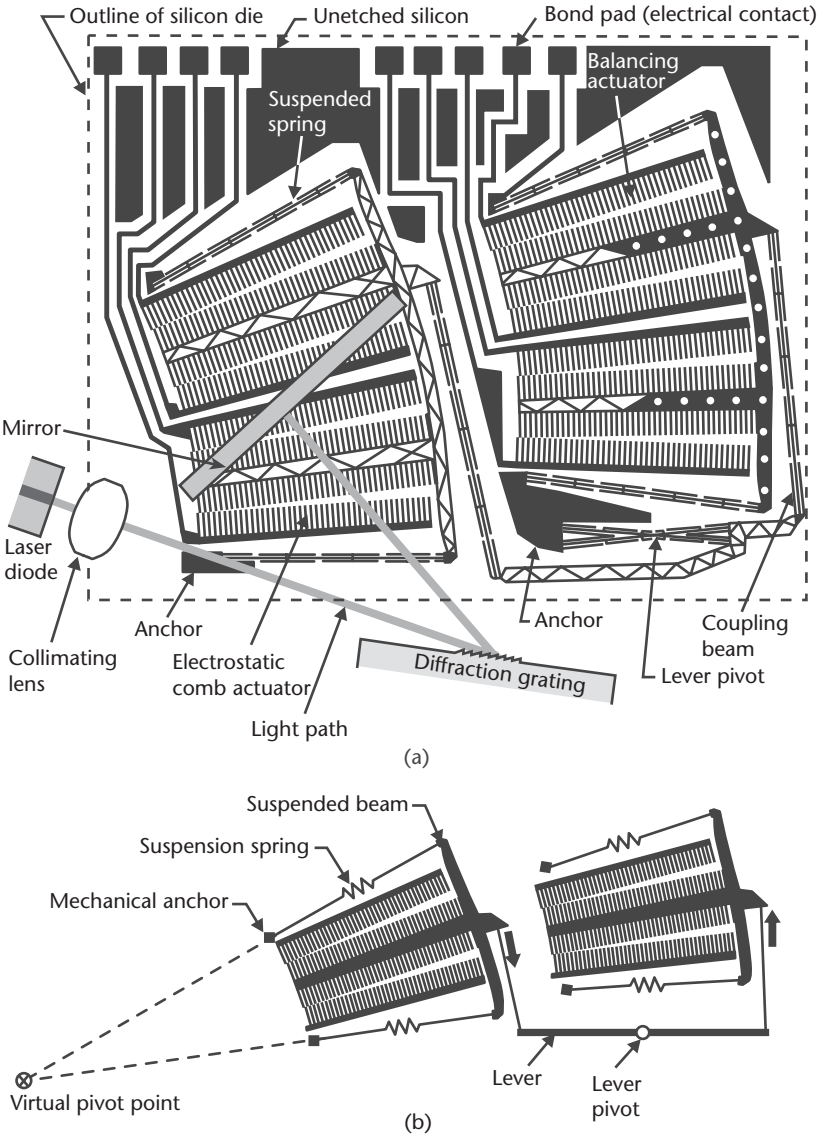


Figure 5.9 (a) Illustration of the mechanically balanced electrostatic comb actuator design with the reflecting mirror. The laser diode, collimating lens, and diffraction are also shown in reference to the actuator. (b) A simplified schematic of the mechanical structure of the comb actuator. A lever in a push-pull configuration connects two comb actuators. The virtual pivot point lies at the intersection of the two flexural suspension beams supporting the loaded actuator with the mirror.

mounting pads forms the reflective mirror end of the cavity. The mirror is 1.7 mm wide and extends approximately 600 μm above the surface of the actuator. The maximum range of rotation of the actuator necessary to tune the laser over the entire C-Band depends on the dispersion of the grating. At 1,200 lines per millimeter, one degree of angular rotation at the mirror causes a 7.5-nm shift in wavelength. Hence, the total required rotation of the actuator is less than five degrees. At this angle, the distal end of the mirror travels 300 μm.

Fabricated using the SFB-DRIE process introduced in Chapter 3, the rotary actuator [18] utilizes a mechanically balanced comb structure with a flexural suspension design [see Figure 5.9(a)]. Its single-crystal silicon design makes it inherently free of intrinsic stresses and hysteretic mechanical effects. With a typical spring width of 4 μm and a thickness of 85 μm, the out-of-plane stiffness is sufficiently high to confine all displacements to the plane of the silicon die. The comb elements are also 4 μm wide with a gap of 10 μm. The fundamental in-plane mechanical resonant frequency is 212 Hz. All flexures and springs include fin-like structures to simulate a periodic structure during the DRIE step, thus minimizing the loading effect (see Chapter 3) and improving the sidewall profile. As these fins are attached only to the suspended flexures and springs, they have no impact on the spring constants, but they add mass and cause a slight reduction in the mechanical resonant frequency.

The rather large thickness and size of the silicon comb actuator result in a relatively high mass that makes the device sensitive to in-plane vibrations and accelerations—an unbalanced actuator behaves similar to the DRIE accelerometer described in the previous chapter. This undesired vibration sensitivity is greatly reduced by a mechanically balanced design that incorporates two electrostatic comb actuators coupled together by a lever in a push-pull configuration [see Figure 5.9(b)]—when one actuator rotates in a clockwise direction, the other turns in the opposite orientation. The combs are nearly identical, differing only in their masses: the mass of the unloaded actuator on the right-hand side is equal to the mass of the loaded actuator (left-hand side) and the mirror. Externally applied in-plane accelerations cause equal but opposite torques on the lever, thus minimizing any undesired motion of the mirror. Nonetheless, minute imbalances between the masses of the two actuators remain and adversely impact the optical length of the cavity. An electronic feedback servo loop monitoring the output wavelength (see the following section on wavelength lockers) applies a force-balancing voltage to the comb structure and counteracts small parasitic displacements, thus eliminating any residual rotation of the mirror. With the servo loop active, the measured optical wavelength shift at an applied sinusoidal vibration of 5G at 50 Hz is less than 10 pm (equivalent to an optical frequency shift of 1.25 GHz off the main optical carrier on the ITU grid at approximately 194 THz).

The orientation of the flexural springs that support the loaded actuator on the left-hand side determines the location of the virtual pivot point. For nonintersecting flexures and small deflections, the pivot point lies at the intersection of the lines extending from these flexures [19]. This design was preferred by the engineers over centrally symmetrical rotary actuators that are inherently balanced because of space considerations in the miniature laser package.

The theory of conventional electrostatic comb actuators teaches that the attractive force is quadratic with the applied voltage [20]. This nonlinear dependence

makes the design of closed-loop electronic circuits rather complex. Instead, it is desirable to design a mechanical system whose force (and hence angular displacement) is linear with applied voltage. A close examination of the actuator reveals that the length of the individual comb teeth varies, becoming shorter towards the outer periphery of the rotary actuator. As the two comb actuators in the push-pull configuration are driven differentially and rotate in opposite directions, additional teeth engage in one actuator and disengage in the other [18]. The rate at which the total number of engaged teeth changes with angle of rotation (and applied voltage) is determined by the geometry and layout of the comb teeth. If the total number of engaged teeth is inversely proportional to the square of the voltage, then the nonlinear dependence is eliminated. In practice, this dynamic tailoring of the number of engaged comb teeth with angle greatly reduces the overall nonlinear dependence but does not eliminate it. Experimental analysis shows that the behavior is generally linear with high-order ripples [20]. For the particular design used by Iolon, a differential voltage drive of 150V results in an angular rotation of ±2.5°.

Once packaged in a standard 18-pin butterfly package (see Chapter 8) with all of the components optically aligned, the product meets all of the requirements of a tunable laser for long-distance transmission. The power is 13 ± 0.1 dBm from 1,529 to 1,561 nm; the RIN measures –145 dB/Hz from 10 MHz to 22 GHz; the SMSR is 55 dB; and the spectral linewidth, typical of external cavity lasers with long cavities, is very narrow, measuring 2 MHz [21]. The tuning speed of the laser is only limited by the actuator's mechanical response time and the bandwidth of the closed-loop servo. A maximum tuning speed of approximately 10 ms has been reported.

The DFB Tunable Laser from Santur Corporation

The laser design used by Santur [11] bears no resemblance to the previous design, other than achieving a similar performance. It is based on a family of integrated semiconductor lasers called distributed-feedback (DFB) lasers [22]. These lasers are ubiquitous as transmission sources in fiber-optic telecommunications owing to their excellent spectral performance and proven reliability. They have been manufactured in volume for many years and thus are cost effective. They provide a stable output power, typically between 10 and 50 mW, and are frequency stabilized by a Bragg grating guaranteeing no wavelength drift. It is common to obtain in a communication-grade DFB a RIN of better than –145 dB/Hz from 50 kHz to 2.5 GHz, a SMSR higher than 45 dB, and a spectral linewidth narrower than 2 MHz (e.g., [23]).

The details of the DFB laser are beyond the scope of this book and can be found in [24]. In summary, the basic structure consists of a gain medium made of multiple quantum wells in an InGaAsP/InP semiconductor crystal (see Figure 5.10). Light is confined within the crystal to a waveguide that is made by the difference in index of refraction between InP and InGaAsP. A periodic Bragg grating delineated immediately above the waveguide provides a wavelength filter as well as a resonant cavity. The Bragg grating reflects light continuously over its entire length, thus behaving as a distributed reflector and resulting in a distributed resonant cavity—hence the name distributed feedback. This coupled role of the Bragg grating makes a full analysis numerically complex and intensive. The grating shape, periodicity, and index of refraction determine the center wavelength of the filter, as well as its

reflectivity into the cavity. If Λ is the periodicity of a simple grating and n' is the difference in indices of refraction of the materials bounding the grating, then the center wavelength of the grating in free space is $\Lambda/2n'$ [25]. For a grating centered at 1,550 nm in InP and InGaAsP ($n' \approx 0.2$), the required periodicity is approximately 0.6 μm, necessitating fabrication using high-resolution lithographic tools such as an electron beam. The dependence of optical gain and index of refraction on temperature results in the lasing wavelength increasing with temperature at the rate of 0.12 nm/°C over the range 20° to 80°C. This is why semiconductor lasers include a TEC device to control temperature and wavelength.

The Santur laser utilizes temperature as the variable parameter to tune the output wavelength of the DFB laser. However, a 25°C change in temperature results in a 3-nm wavelength shift that is only a fraction of the entire C-Band. This limited thermal tuning range gives rise to using a linear array of 12 DFB lasers. All are similar in every respect, differing only in the periodicity of their Bragg gratings, each covering a small portion of the C-Band (about 3 nm) (see Figure 5.10) [26]. Applying a current to a particular laser in the array selects this laser for operation; a temperature adjustment then fine tunes the output wavelength. A tilting micromachined mirror then steers the output light beam through a focusing lens into an optical fiber. The micromirror only needs to tilt in one direction for laser selection, but a tilt capability in the orthogonal direction aids in relaxing the alignment tolerances during final packaging. The maximum angular tilt is quite small, only about ±1.5°, because the DFB lasers in the array are on a 10-μm pitch.

Unlike the external cavity laser described earlier, this laser resonant cavity is fully contained within the semiconductor diode, and, hence, external vibrations have no effect on the output wavelength. However, these external vibrations may cause minute misalignments of the micromirror relative to the two lenses, thereby modulating the output power coupled into the optical fiber. Experimental

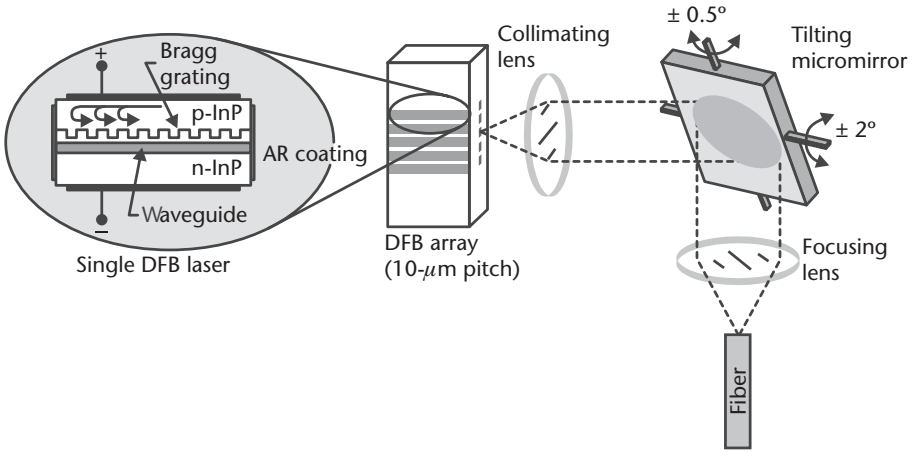


Figure 5.10 Schematic illustration of the tunable array of DFB lasers from Santur Corporation of Fremont, California. Once a DFB laser in the array is electrically selected, a micromirror steers its output light through a focusing lens into an optical fiber. Changing the temperature of the DFB laser array using a TEC device tunes the wavelength over a narrow range. The illustration on the far left depicts the simplified internal structure of a single DFB laser. Both facets of the semiconductor diode are coated with an antireflection (AR) coating.

measurements showed that a 10-G shock without compensation by the electronic control loop caused a small but noticeable 0.2 dB change in output power.

A fully packaged laser contains three electronic control closed-loop circuitries that use input from a commercial wavelength locker to measure output wavelength from a 1% optical tap to measure output power in the fiber [11]. The first circuit controls the drive current to operate one single DFB laser in the array and compensate for any change in the output intensity resulting from long term aging effects. A second circuit controls the temperature of the DFB laser array, thus tuning the output wavelength. The third circuit applies the appropriate voltages to the micromirror to maximize optical coupling and offset any mechanical misalignment. An intentional misalignment of the micromirror angle can attenuate the coupled power into the fiber and turn the mirror into an integrated variable optical attenuator. The present design offers a 10-dB attenuation range, thus providing an output power that can be varied from 3 dBm (2 mW) up to 13 dBm (20 mW). It is evident that when the micromirror is in an extreme angular position, no light couples into the fiber, thus *blanking* the laser. This is necessary during the transient duration when the control electronics are switching between different DFB lasers within the array or altering the temperature of a single array element.

The final measured specifications of this tunable laser are as follows: the maximum output power is 13 ± 0.1 dBm tunable from 1,531 to 1,564 nm; the RIN is lower than –145 dB/Hz up to 10 GHz; the SMSR is 43 dB; and the spectral linewidth is 8 MHz. The tuning speed of the laser is limited by the response of the TEC to about 1s. We will discuss the packaging details of this tunable laser in Chapter 8.

The micromirror is a double-gimbaled structure (see Figure 5.11) fabricated using bulk-micromachining methods in single-crystal silicon. The mirror is a highly doped, conductive silicon layer that is a few micrometers thick and is suspended using torsional flexures from a frame. The inner and outer torsion flexures are 3.5 μm and 5 μm wide, respectively; both are 150 μm long and 15 μm thick. The

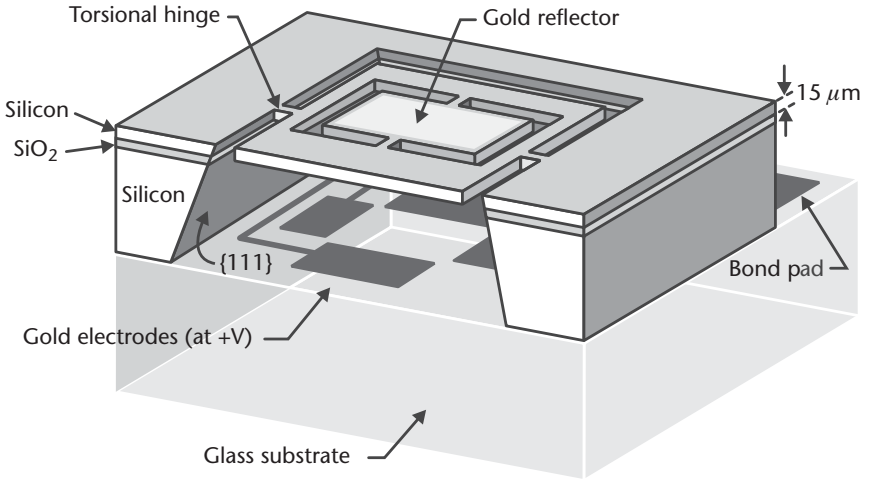


Figure 5.11 Schematic cross section of the micromirror used within the tunable laser from Santur Corporation. The device consists of a double-gimbaled mirror structure supported by torsional hinges. A gold layer defines the high-reflectivity mirror surface that remains at ground potential. Four gold electrodes on an anodically bonded glass substrate actuate the mirror and cause rotation around the hinges.

reflective element is a layer of gold measuring 1.4 × 1.5 mm^2. The large but finite reflectivity (~ 98%) of gold results in some coupling losses measuring about 0.2 dB. Four gold electrodes on a glass substrate directly underneath the mirror actuate the mirror and cause rotation. The dimensions of the torsional flexures and electrode-mirror gap are such that full angular rotation is obtained at about 120V. The overall die measures 3 × 3 mm^2.

Santur Corporation has disclosed sufficient details to recreate a fabrication process. One starts with a SOI wafer that is approximately 155 μm thick composed of a crystalline top silicon layer (15 μm thick) on top of a 0.25-μm thick silicon oxide layer over a silicon handle substrate. A gold layer (20 to 50 nm thick) is deposited either using evaporation or sputtering. Because gold deposited in either method tends to be under stress, it is desirable for the SOI silicon layer to be as thick as possible. The torsion flexures and mirrors are then lithographically delineated and etched into the SOI silicon layer using standard dry etch methods. The front side is protected, and the bulk of the wafer is etched from the back side using KOH or TMAH. The etch stops at the intermediate silicon oxide layer. An etch of the oxide layer then releases the suspended mirror structure. A flash sputter deposition or evaporation of gold on the back side of the mirror can greatly alleviate any bowing of the mirror due to stresses from the gold film on the front side by adding a balanced stress on the back side. Finally, the silicon wafer is anodically bonded to a glass substrate upon which gold electrodes were previously deposited and lithographically delineated.

Wavelength Locker

The function of a wavelength locker is to measure the difference between the actual and desired wavelengths of a laser and provide an error signal that can be used by the laser's electronic control loop to correct for the deviation. This function is key to all lasers used in fiber-optic telecommunication in order to "lock" the laser output to an assigned wavelength on the ITU grid and to offset drift due to aging and environmental conditions. Wavelength lockers are available as stand-alone products external to the laser or can be integrated within the laser resonant cavity. This section selects for description a micromachined wavelength locker from Digital Optics Corporation of Charlotte, North Carolina, which operates in the wavelength range from 1,525 to 1,625 nm.

A wavelength locker is a special implementation of a one-dimensional planar resonant cavity known as a Fabry-Perot etalon [27]. For an etalon of physical length d and index of refraction n bounded by two partially reflective surfaces of reflectivity r, the frequency spacing between adjacent longitudinal modes is known as the free spectral range (FSR) and equals $c/2d$ under normal incidence. The transmission transfer function of the etalon, therefore, consists of periodic peaks whose sharpness is a function of the *finesse*, a measure of the loss in the resonator's mirrors. A low mirror reflectivity results in a low finesse and broad transmission peaks [see Figure 5.12(a)]. The utility of the etalon is that once its transmission peaks are calibrated, it forms a wavelength "ruler" against which an incident beam of unknown wavelength can be measured. The calibration process entails fixing the frequency position of the transmission peak relative to the ITU grid and ensuring that the FSR matches the periodicity of the ITU grid, typically 50 GHz. An etalon is a simple structure, but to obtain a precise and accurate measurement of the

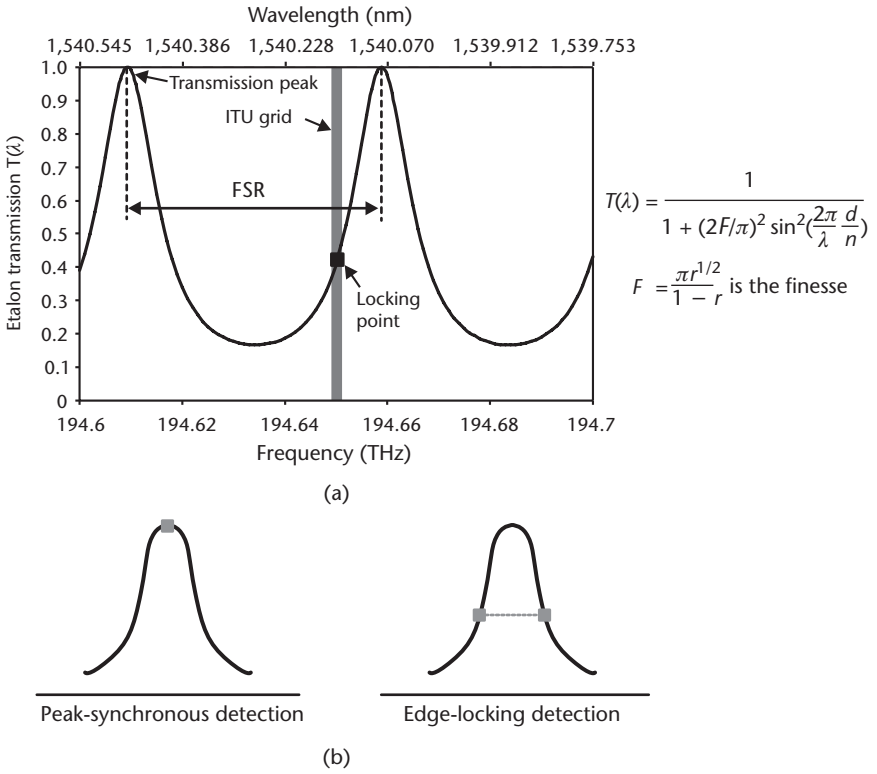


Figure 5.12 (a) The transmission function of an example etalon in the infrared spectrum showing the transmission peaks relative to the ITU grid. When calibrated, the FSR matches the periodicity of the ITU grid. The transmission peaks become narrower and sharper with increasing finesse (higher mirror reflectivity). (b) The laser can be locked to the etalon's transmission function using peak-synchronous detection (locking to the peak) or edge-locking detection (locking to a predetermined intensity level).

frequency (or wavelength), its transmission peaks (equivalent to the marks on a ruler) must be calibrated and insensitive to temperature and process variations. It is the fabrication of an etalon that is complex and requires great diligence.

A basic wavelength locker consists of a beam splitter, a Fabry-Perot etalon, and two detectors. An incident light beam is divided into two optical beams. The first one travels through a calibrated etalon, and its intensity is measured by a first detector. The second beam is directly detected by a second detector and serves as an intensity reference. The differential analysis serves to eliminate the effect of any power fluctuations in the laser diode itself on the wavelength measurement. The transmission function of the etalon maps any wavelength changes in the incident beam to an intensity change that is measured by the detectors. An electronic control loop takes the differential error signal between the two detectors as input and closes the loop to the laser control circuitry to lock its output wavelength to a predetermined value. Depending on the design of the electronic circuitry, the locking can be peak synchronous (locking to the transmission peaks) or edge level (locking to a predetermined intensity level) [see Figure 5.12(b)]. The latter locking method is common because the high slope at the locking point results in a large change in intensity for a small change in wavelength (or frequency), typically 2% to 7% per GHz [28].

The wavelength locker from Digital Optics Corporation miniaturizes the elements of a standard locker to result in a device that measures 5.5 × 5.5 × 4.0 mm^3 (see Figure 5.13). The miniature beam splitter is a diffractive optical element consisting of a multiple-phase grating [29, 30]. Relief structures etched in an optically transparent substrate, such as fused silica or silicon in the infrared portion of the spectrum, form regions of varying phases and thus diffract the incident light in a specific pattern or orientation. The efficiency of the grating (the intensity that is available in the higher order diffraction orders) is tailored by adjusting the planar topography and phase of the relief structures. A thorough analysis typically involves using a fast Fourier transform of the output light intensity to calculate the phase pattern of the diffractive element. In this particular case, a high-efficiency phase grating divides the incident light beam into two higher order beams noted as +1 and –1 orders. A second set of similar diffractive optics corrects the path of the two diffractive beams (and possibly collimates them) for normal incidence on the detecting photodiodes.

The etalon is a cube of fused silica (n = 1.444) whose polished facets have been coated with a partially reflective dielectric mirror. For the wavelength range of 1,550 to 1,650 nm, the dielectric mirror typically consists of alternating layer pairs, each one quarter of a wavelength in thickness, of Al_2O_3 and SiO_2, or TiO_2 and SiO_2. The optical thickness of the glass cube determines the free spectral range. For the locker to work properly on the ITU grid, the FSR must equal the ITU channel separation of 50 GHz ± 0.01 GHz. This means that the smallest thickness of the etalon must equal to approximately 2.076 mm. The fine calibration of the FSR to the 50-GHz ITU spacing relies on the dependence of the index of refraction, n, on temperature to adjust the optical length of the etalon. For fused silica, the temperature coefficient of

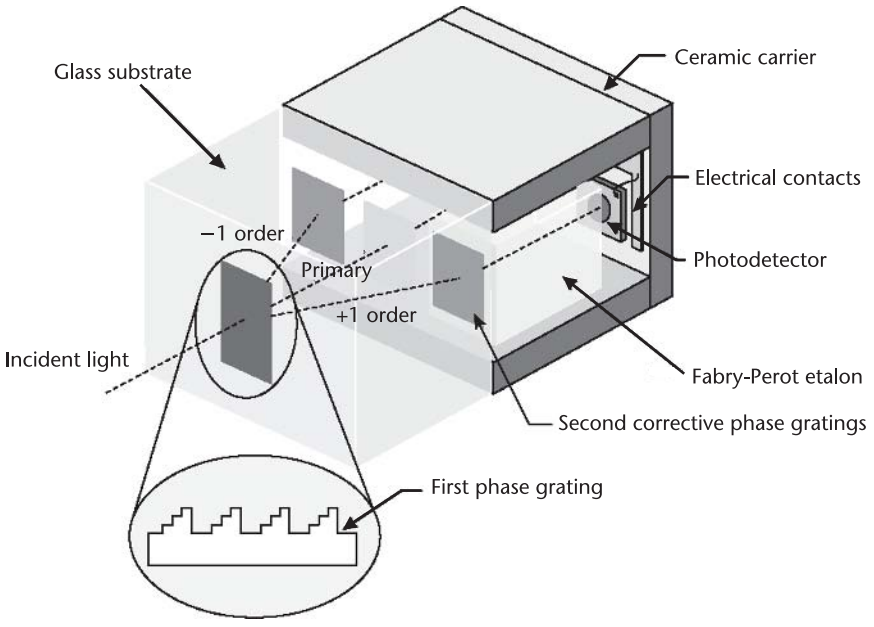


Figure 5.13 A schematic illustration of the wavelength locker from Digital Optics Corporation. A first diffractive element divides the light from the laser into two beams. The first one travels through a calibrated etalon and is detected by a photodetector. The second is detected directly by another photodetector and serves as an intensity reference.

its index of refraction, dn/dT, is approximately $7 \times 10^{-6}K^{-1}$ at 1,545 nm. This temperature dependence results in a spectral sensitivity of 1.35 GHz/°C, thus necessitating the use of a TEC to stabilize the etalon's temperature.

The full product is a subassembly that comprises a first glass substrate with the diffractive optical elements, an etalon on a ceramic carrier, and two photodetectors also mounted with the appropriate electrical contacts on a ceramic carrier. The subassembly elements are epoxied together and cured. The optical path is usually void of any epoxy to ensure long-term reliability.

Digital $M \times N$ Optical Switch

An optical switch is to fiber-optical telecommunications and light transmission what a multithrow, multipole electrical switch is to electronic or microwave signaling. An optical switch redirects an incoming light signal into one of many output fibers. Arrays of optical switches allow the rapid reconfiguration of optical networks in data communications by altering the light path in a system of intersecting fibers—much like railroad points move train tracks to reconfigure a rail network. They are also useful components in the addition and deletion of extra channels in optical add/drop multiplexers (OADM) [31] for WDM—equivalent to highway entrance and exit ramps that add or remove some of the traffic. During the boom years of the late 1990s, venture capital funded a large number of innovative startup companies in a race against the giants of the telecommunications industry to develop the next generation of optical switches. But the collapse of the telecommunications market in the following years resulted in the closure of many of these companies and forced the remaining ones to demonstrate products that were cost effective and highly reliable.

A key characteristic of optical switch arrays is their order (i.e., how many input and output fibers can be independently coupled to each other). If a switch can route the light from a single input fiber to any of N output fibers, then it is labeled $1 \times N$. Generally, $M \times N$ switches are two-dimensional arrays with M input and N output fibers. Their electronic equivalent is an analog multiplexer that selects any one of M electrical inputs and routes its signal to any one of N output lines. The now-defunct company Optical Micromachines, Inc., formerly of San Diego, California, was one of the early pioneers to demonstrate arrays using surface-micromachined polysilicon mirrors in switches reaching up to 32×32. Optical switches are commercially available today from such companies as JDS Uniphase of San Jose, California; DiCon Fiberoptics, Inc., of Berkeley, California; and Sercalo Microtechnology, Ltd., of Liechtenstein but are typically 1×2 up to 1×8, or they are 2×2 switches.

The basic architecture and technology used by Sercalo for their $M \times N$ switches stems from prior research activities performed at the University of Neuchâtel in Switzerland using the SOI and DRIE process introduced in Chapter 3 [32]. While not all of the details of the Sercalo switch are publicly available, the device is nearly identical to the work completed at the University of Neuchâtel, and we will thus limit this discussion to the latter device.

The basic cell for a 2×2 switch element consists of an electrostatic comb actuator controlling the position of a vertical mirror plate at the intersection of two perpendicular slots. Within each slot lie two optical fibers, one on each slot end (see Figure 5.14). In the actuator's normal unbiased position, the mirror plate sits in the

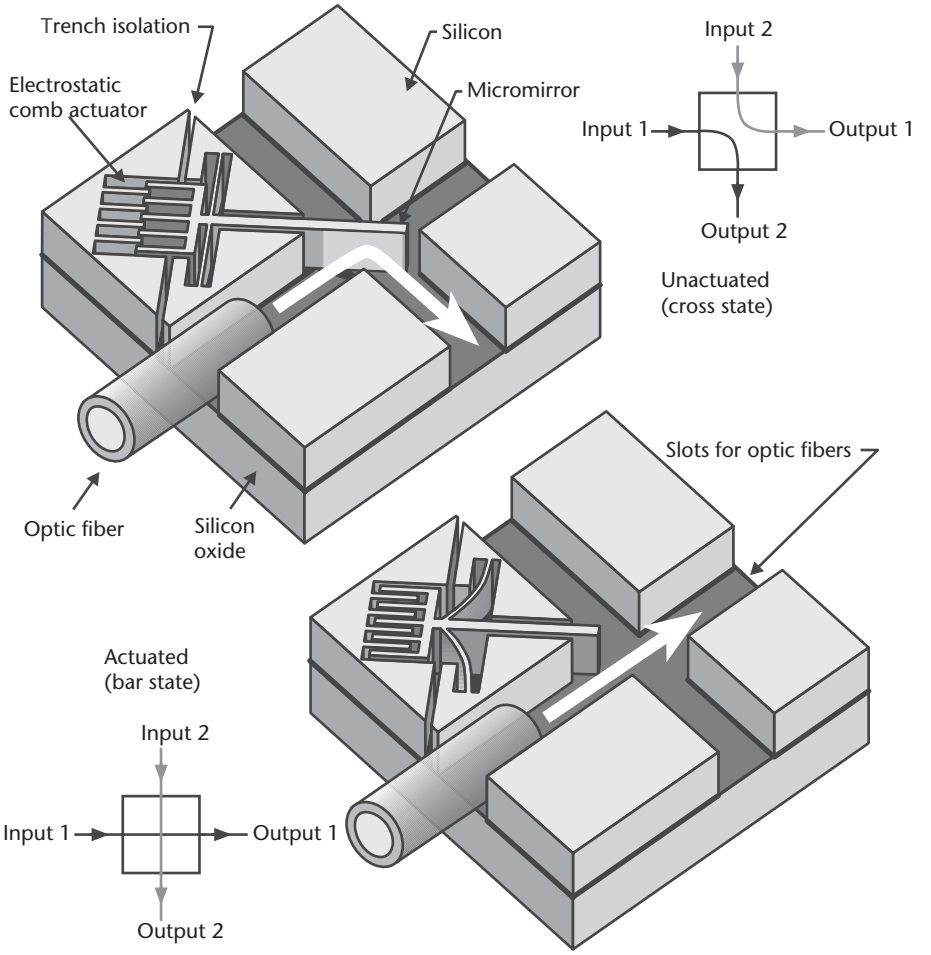


Figure 5.14 Illustration of a 2 × 2 binary reflective optical switch fabricated using SOI wafers and DRIE. An electrostatic comb actuator controls the position of a micromirror. In the cross state, light from an input fiber is deflected by 90°. In the bar state, the light from that fiber travels unobstructed through the switch. Side schematics illustrate the signal path for each state.

middle of the intersection and reflects the light by 90°, thereby altering the path of data communication—this is the cross state. Applying approximately 70V to the actuator combs causes the mirror to retract, letting the light pass through unobstructed—this is the bar state. Arraying the 2 × 2 switch element in both directions creates a generalized $M \times N$ switch matrix.

The slots must normally accommodate optical fibers, typically 150 to 250 μm in diameter. The depth of the slot must be such that the center of the fiber aligns with the center of the micromirror. Because light diverges upon leaving the fiber, the mirror must be significantly larger than the fiber core itself (the core is a central area that carries light, typically about 10 μm in diameter for single-mode fibers [33]). In the demonstration from the University of Neuchâtel, the mirror height is identical to the depth of the groove, approximately 75 μm. Insertion loss, a measure of the light-coupling efficiency between input and output fibers—it is the ratio of intensities at the output to the input—greatly depends on the alignment accuracy

of the fibers with respect to each other and to the mirror. Insertion loss also relies on the mirror reflecting all of the light impinging on its surface. This essentially requires the use of highly reflective coatings, in particular aluminum for wavelengths in the visible and gold in the infrared. Furthermore, the surface of the mirror must be optically flat in order to eliminate any deleterious light-scattering effects. A surface roughness less than 10% of a wavelength is considered to be optically flat for most applications.

The device is fabricated on SOI wafers with a 75-μm-thick top silicon layer. The overall die size is approximately 3.3×3.7 mm^2. Lithography in standard resist was followed by DRIE down to the buried oxide. An etch step in hydrofluoric acid removes the 2-μm buried silicon dioxide layer to release the comb actuator as well as the mirror plate. In the research performed at the University of Neuchâtel, 50-nm-thick aluminum is deposited on the silicon surfaces to increase reflectivity. Owing to its small size, small mass, and stiff springs, the switch has a fast response time, typically 500 μs.

The optical switch from the University of Neuchâtel demonstrated an overall insertion loss of less than 1.6 dB in the bar state and less than 3.4 dB in the cross state, losses considered excessive for optical-fiber telecommunications. The measured reflectivity of the aluminum mirror was 76%, lower than the theoretical value of 95%, thus contributing about 1.2 dB to the loss ratio [34]. The loss was further exacerbated by scattering from the surface roughness of the mirror plate, which was measured at 36 nm rms.

In producing a commercial version of the switch, Sercalo had to address these problems to offer adequate performance in the operating wavelength range from 1,300 to 1,600 nm. First, substituting gold for aluminum greatly improves reflectivity in the infrared. Second, developing a gold deposition process that yields a smoother, fine-grain structure is necessary to reduce scattering effects. Sputtering, evaporation, and slow (low-current) electroplating of gold are known to yield smooth surfaces. Finally, the fibers must be aligned relative to each other with an accuracy of less than 0.5 μm to minimize coupling losses. Because all of the slots and mirror are precisely defined in a single lithographic and etch step, any misalignment will typically result from variations in the fiber diameter, which can be greatly minimized by using fiber from the same spool. As a result of these improvements, Sercalo was successful in reducing the insertion loss to an average of 0.5 dB.

The measured polarization-dependent loss (PDL), which is the ratio of reflected intensities corresponding to the two polarizations, is less than 0.1 dB. PDL is primarily affected by the dependence of the reflectivity of metallic surfaces on polarization, especially at oblique incidence. For a 45° incident angle on a gold surface, the difference in reflectivity is about 0.5%, which corresponds to a contribution of 0.025 dB to PDL.

Beam-Steering Micromirror for Photonic Switches and Cross Connects

In $M \times N$ switch arrays, the number of individual mirrors is $M{\cdot}N$. Arrays bigger than 32×32 require a substantially large number of mirrors and become subject to sufficiently low manufacturing yields, which make them uneconomical. Furthermore, the insertion loss increases with the array size as the optical path length gets longer, and sensitivity to misalignment becomes more critical. A new

three-dimensional (3-D) system architecture using continuously tilting mirrors in two directions can serve the same functionality but with far fewer mirrors: an optical cross connect with N input and output fiber ports requires only $2N$ mirrors. The mirrors in the 3-D architecture are no longer digital (*ON-OFF*), but rather point the light beam from one fiber to another with high spatial precision (see Figure 5.15). In this beam-steering approach, a first tilting mirror on a first plate points the light from a collimated input fiber to one of many similar mirrors on a second plate, which in turn points the light to a collimated output fiber. Both an input and an output mirror are required, so that each can be pointed directly at the centerline of its corresponding fiber, rather than at an angle. To minimize the maximum angular displacement of the mirrors, the two plates can be positioned at 45° relative to the incident light. The angular tilt precision needs be very high. For a system using single-mode fibers with a typical core diameter of 10 μm and an optical path length of, say, 10 cm, the mirror must have an accuracy and repeatability of better than 100 μrad (0.006°). The system specifications require a mirror design that is capable

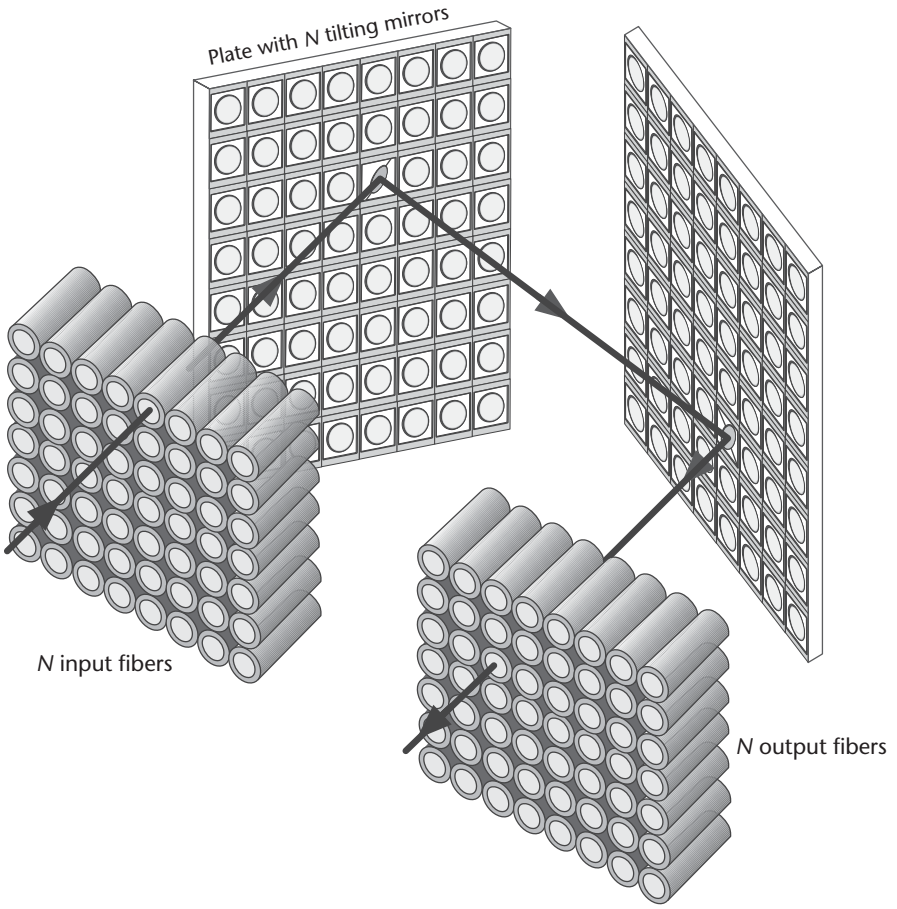


Figure 5.15 Schematic illustration of the 3-D architecture for an $N \times N$ switch or photonic cross connect. A beam-steering micromirror on a first plate points the light from a collimated input fiber to another similar micromirror on a second plate, which in turn points it to a collimated output fiber. This system architecture requires a total of $2N$ continuously tilting mirrors in two directions. To minimize the maximum angular displacement of the mirrors, the two plates are positioned at 45° relative to the incident light.

of delivering such high angular pointing precision, as well complex alignment and closed-loop feedback systems that are beyond the scope of this book [35].

Many companies achieved significant progress in the development of very-large-scale photonic switches (reaching up to 4,000 × 4,000). However the collapse of the telecommunications capital equipment market by 2002 forced many of these companies to use their micromirror technologies in pursuit of other, possibly less lucrative, markets. It is the broad utility of these beam-steering micromirrors that leads us to present them here, even though the primary application for which they were developed (fiber-optical telecommunication) will not witness significant growth until a future time.

A search on issued patents in this field reveals a plurality of micromirror inventions, the vast majority of which utilize electrostatic actuation (e.g., [36]). One implementation from Integrated Micromachines, Inc. (IMMI), of Irwindale, California, utilizes electromagnetic actuation instead. While the company is no longer pursuing applications in fiber-optical communications, the design stands out as an elegant implementation using a low-voltage, low-power electromagnetic scheme [37].

The basic design for virtually all beam-steering micromirrors, including the device from IMMI, consists of a bulk-micromachined mirror supported from a silicon frame using a gimbal suspension (see Figure 5.16). The mirror is often circular in shape, though elliptical, rectangular, and square shapes are also possible. A thin layer (10 ~ 100 nm) of metal on the surface ensures a high reflectivity; gold is the metal of choice for infrared radiation. The IMMI design utilizes four independent drive coils on the back side of mirror for actuation.

The design places the reflective surface of the mirror on what conventionally is the back side of a double-polished silicon surface. The thickness of the mirror is approximately 100 to 200 μm. A thick and thus stiff mirror is essential to reduce the risk of distortions (e.g., warping due to heating from absorbed laser radiation or stress from the deposited gold layer). The mirror diameter is often 5 to 10 times the nominal diameter (typically measured as full width at half maximum) of the light beam to ensure that the mirror intercepts *all* incident rays. This in itself assumes that

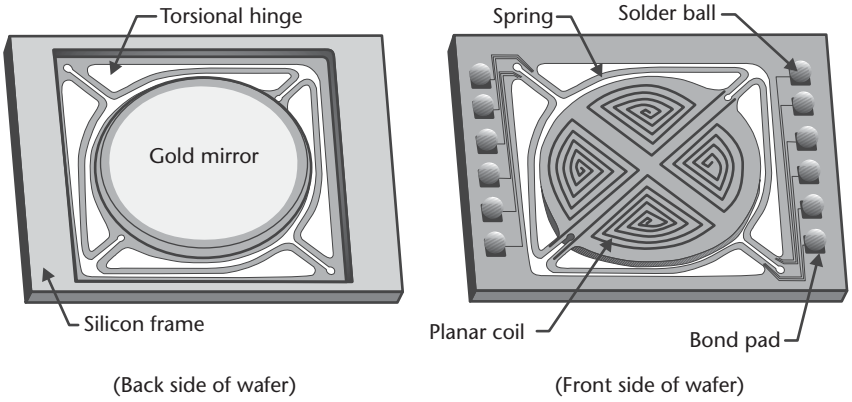


Figure 5.16 Schematic illustration of the beam-steering micromirror from IMMI. The mirror is formed on what conventionally is the back side of a double-polished SOI wafer, while the drive coils and thin torsional flexures are made on the front side. Sn-Pb solder balls allow the packaging of arrays of mirrors on ceramic substrates using flip-chip technology.

the propagating radiation is in a single Gaussian mode—multimode radiation will require much larger diameter multiples [38]. The IMMI design micromirror has a diameter of approximately 3 mm, giving it a mass of 1.7 mg.

The gimbal suspension consists of four serpentine torsional hinges arranged in a symmetrical topography and formed in the top silicon layer of what conventionally is the front side of a SOI wafer. This allows the manufacture of thin, compliant hinges, which results in lower actuation forces [39]. However, if the hinges are too compliant, the suspension-mirror mechanical system will be sensitive to vibration and will not survive mechanical shocks. The final dimensions are thus a compromise depending on many factors, including the magnitude of available actuation forces, required size of the mirror, available real estate, and allowed resonant modes. The suspension-mirror geometry and dimensions are such that the first resonance of the IMMI mirror is at 140 Hz. The present gimbal suspension favors three modes of displacement (two out-of-plane angular rotations and one out-of-plane displacement), but it also permits additional undesirable modes such as in-plane motion or rotation of the mirror. Fortunately, these undesirable modes have resonant peaks above 3 kHz and thus do not participate in the mirror motion, provided the control electronics limit the bandwidth to a value lower than these resonant frequencies. Numerical analysis of the suspensions and experimental results has shown that the rotational spring constants remain unchanged through the full angular displacement of the micromirror. Consequently, the mirror actuation is linear with current in the drive coils, a feature that simplifies the implementation of the control electronics.

Magnetic actuation is a key differentiator of the IMMI micromirror, as it delivers a higher actuation energy per unit volume compared to equivalent electrostatic actuation methods (see Table 4.2). A larger actuation force enables the use of a relatively stiffer suspension and thicker mirror, thus improving the overall mechanical response. The actuation force is given by the Lorentz force and depends on the following key parameters: the length and orientation of the drive coil and the intensity and orientation of the magnetic field vector. The drive coils are formed by electroplating on the front side of the wafer with electrical connections leading to tin-lead (Sn-Pb) solder balls made using standard screen printing and reflow processes. The solder balls allow the packaging of multiple mirrors in arrays on ceramic substrates using flip-chip technology (see Chapter 8).

There are a total of four coils, one in each quadrant of the circular mirror. The coils reside within a short distance (200~500 μm) from the surface of a permanent rare-Earth cylindrical magnet. The magnetic flux density at the surface of the magnet is approximately 1T but rapidly decays with distance. The magnetic flux density outside of the magnet has two components: normal ($\mathbf{B}_n$) and radial ($\mathbf{B}_r$) [see Figure 5.17(a)]. The total actuation force consists of the contributions of both components to the Lorentz force. A counterclockwise current interacting with the normal component $\mathbf{B}_n$ results in a Lorentz force that acts in the plane of the coil [see Figure 5.17(b)]. $\mathbf{B}_n$ is not constant across a coil, resulting in a net force that is radially outward for a single coil. By pairing the coils in a symmetrical manner, the in-plane forces from all four coils counteract each other, thus greatly reducing motions in the plane of the mirror. A suspension with high in-plane stiffness further ensures that in-plane motion is negligible.

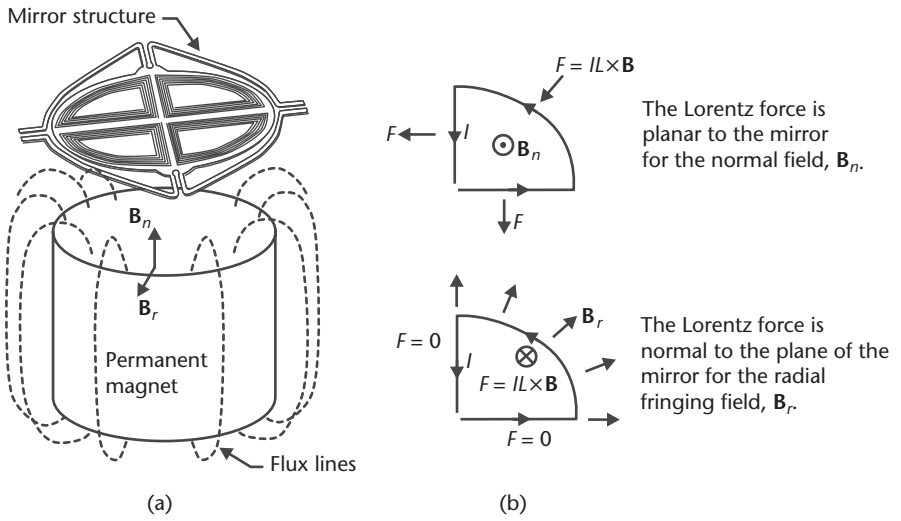


Figure 5.17 (a) An illustration of the rare-Earth magnet and the four independent drive coils. The magnetic flux density outside of the magnet has a normal component, $\mathbf{B}_n$, and a fringing radial component, $\mathbf{B}_r$. (b) The normal magnetic component interacts with a counterclockwise current to induce a Lorentz force that is in the plane of the coils. The radial component of the magnetic field results in a force that is normal to the plane of the coil.

Only the fringing radial component of magnetic field ($\mathbf{B}_r$) contributes to the angular and piston (vertical) displacement of the micromirror. A counterclockwise current in a drive coil interacting with the radial field results in a Lorentz force that is normal to the plane of the coil and acting to pull the coil towards the magnet [see Figure 5.17(b)]—the peripheral portion of the coil contributes to the force, whereas the radial portions have little effect. Switching the polarity of the current results in an opposite force that pushes the coil away from the magnet. It thus becomes evident that two adjacent coils carrying currents in opposite directions induce a torque around an axis of symmetry that divides them. Torques of arbitrary magnitude can be generated around the two axes of symmetry by the proper selection of the current direction and magnitude in each of the coils. Furthermore, an additional vertical (piston) motion can be induced by driving all four coils simultaneously with a current in the same direction. For example, a clockwise current in all coils moves the mirror away from the surface of the magnet.

The differential drive of the coils provides an added benefit: the developed torque stays relatively constant throughout the full range of motion of ±5°. As the mirror tilts, the side that is closer to the magnet develops a larger downward force, whereas the side that is farther from the magnet develops a smaller upward force. The two effects are offsetting, resulting in a minimal increase in the torque (<0.2%) over the full mirror travel. This linear behavior greatly minimizes cross coupling between the two axes of rotation (<0.1% in displacement cross coupling).

The drive coils play an additional role as sense coils to detect the angular position of the mirror. A multiturn planar coil deposited on the ceramic substrate that holds the silicon micromirror acts as the primary winding of a transformer, with the four drive coils as the secondary. An ac signal at a frequency of approximately 5 MHz in the primary produces a corresponding sense voltage in each of the four coils

through mutual inductance coupling (the mirror does not respond to this high frequency). This coupling is a strong function of the position and orientation of the coils relative to the primary coil. These sense voltages then become a direct measure of the angular position of the mirror and are used in a closed-loop electronic circuit to spatially lock the mirror.

The details of the fabrication process are not available, but, once again, one can design a fabrication sequence that can produce a similar device. The starting material is a SOI substrate polished on both sides. The first fabrication steps cover the formation of the drive coils and corresponding interconnects on the front side of the SOI wafer. A gold seed layer, typically 50 to 100 nm thick, is sputtered on both sides of the wafer, then followed by standard lithography on the front side to delineate the coil layout. The thin gold layer on the back side will ultimately serve as the reflecting surface of the mirror. Electroplating 5–20 microns of gold on the front side forms the coils and bond pads. The next step is the delineation of the torsional hinges, also on the front side of the wafer. This is completed using standard lithography, followed by standard RIE. It may be necessary to delineate the suspension hinges just prior to the electroplating if the thickness of the gold is more than 5 μm in order to avoid the deposition of resist over the thick topographical features of the gold coils. The fabrication is completed by etching from the back side of the wafer the contour of the mirror and using the embedded silicon oxide layer as an etch stop. Either DRIE or wet anisotropic etching (e.g., KOH or TMAH) can be used. The very last step is the removal of the exposed silicon oxide layer using hydrofluoric acid.

It is evident from this process that the thickness of the suspension is determined by the thickness of the top SOI layer, typically a few micrometers thick. As a result, the mechanical properties of the suspension are very predictable and well controlled. Similarly, the thickness of the mirror is determined by the thickness of the handle layer (thick bottom layer) of the SOI wafer and is uniform—the measured surface flatness over the 3-mm diameter mirror is less than 15 nm RMS with local roughness of approximately 2 nm. The gold layer on the back side of the wafer provides a very high reflectivity in the near infrared spectrum.

Achromatic Variable Optical Attenuation

A variable optical attenuator (VOA) is a dynamic optical component used in fiber-optical telecommunications to adjust the intensity of light inside the fiber. A VOA typically maintains the power below 20 mW, which corresponds to the onset of nonlinear effects such as four-wave mixing, Brillouin scattering, and Raman scattering [40, 41]. Key characteristics of a VOA are spectral range (typically between 1,528 to 1,620 nm), insertion loss (a measure of light lost within the component exclusive of the required attenuation, typically less than 1 dB), polarization-dependent loss (a measure of the difference in loss between the two orthogonal polarizations, typically less than 0.5 dB), wavelength dependence of attenuation (typically less than 0.3 dB over the spectral range), and finally size (a volume less than 1 cm^3 is highly desirable). All loss parameters are measured in dB.

Numerous implementations using MEMS technology have emerged in the past few years. The following example is a product by Lightconnect, Inc., of Newark, California, that utilizes a principle of operation and a structure that are identical to the GLV discussed earlier in this chapter [42]. The basic concept is to use diffraction

to shift energy away (and thus attenuate) from the main undiffracted beam into higher order beams (see Figure 5.18), attenuating the incident beam (attenuation is equivalent to creating a continuum of gray shades). The closely spaced suspended reflective ribbons used for the GLV form the elements of an adjustable-phase grating. When the ribbons are coplanar, incident light is reflected back into the aperture without attenuation. When alternating ribbons are pulled down using electrostatic actuation by one quarter of a wavelength ($\lambda/4$) relative to their adjacent ribbons, the incident energy diffracts into higher orders that are directed outside the aperture, and the incident beam is completely attenuated. When the separation is less than $\lambda/4$, the incident beam is partially attenuated, as some energy is shifted into the higher diffracted orders.

While the VOA derives its basic principle of operation from the GLV, it must also address a number of specifications that are particular to fiber-optical telecommunications. The first one relates to the chromatic dependence of the diffraction grating. Displays have to manipulate only three basic colors: red, green, and blue. But VOAs must manipulate a nearly continuous spectrum of wavelengths from 1,528 nm to 1,610 nm without a chromatic dependence. The second specification is polarization-dependent loss. A difference in attenuation between the two polarizations that is larger than 0.5 dB greatly increases the risk of data errors during transmission. The design from Lightconnect adapts the GLV diffractive technology with two key modifications to applications in fiber-optical telecommunications.

In order to understand the basic operation of the achromatic design, one needs to refer to the use of phasors for time-varying electric fields [43]. In the case of the GLV, two phasors—one for each of the fixed and moveable ribbons—affect the

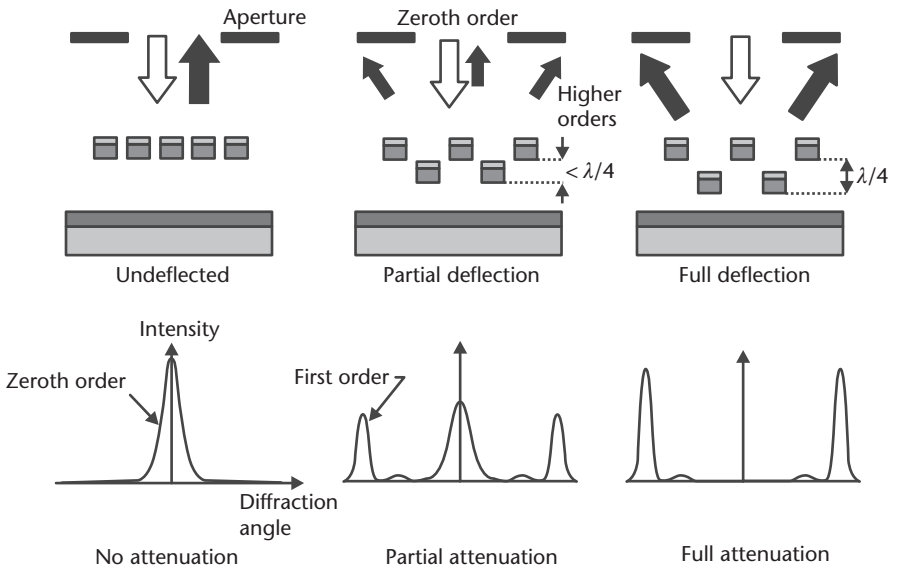
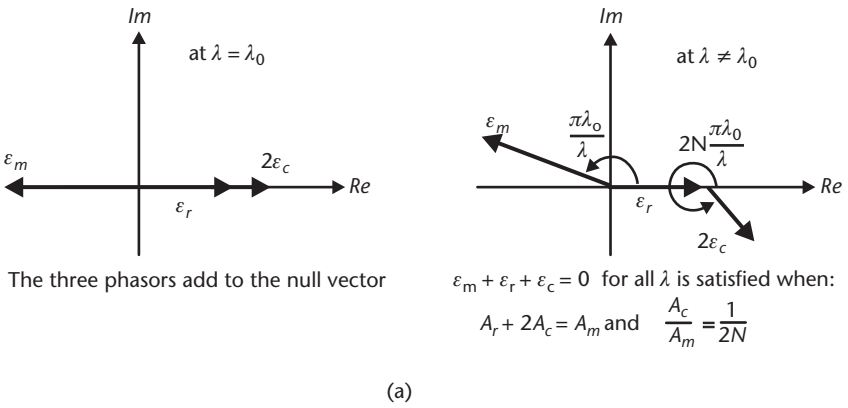


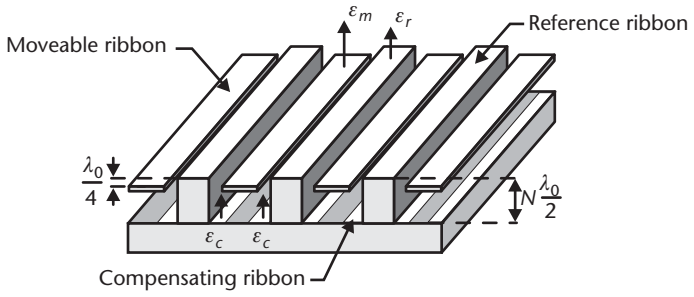
Figure 5.18 An illustration of the basic principle of operation of the variable optical attenuator from Lightconnect, Inc. A set of suspended ribbons act as an adjustable grating. When alternating ribbons are pulled down by $\lambda/4$, the structure becomes a phase grating and diverts the incident energy into higher diffraction orders, thus providing full attenuation of the incident beam. When all of the ribbons are coplanar or separated by a half wavelength, the surface acts as a reflector. When the separation between adjacent ribbons is less than $\lambda/4$, there is light in all orders and the incident beam is only partially attenuated.

reflected wave [see Figure 5.19(a)]. The difference in angle between the two phasors is equal to $4\pi d/\lambda$, where d is the physical separation between the ribbons and λ is the wavelength. When the two phasors are π radians apart (i.e., the total vector sum of the phasors is zero), there is complete diffraction of light into the higher orders. However, this condition is satisfied only at one wavelength, which depends on the separation d. For all other wavelengths, the angle difference between the phasors is less than π (the vector sum is nonzero), thus allowing light to be reflected in both the zeroth (undiffracted) and higher-order diffraction modes. To correct for this dependence, the design introduces another phasor such that the sum of all three vectors is null over a broad range of wavelengths [see Figure 5.19(a)].

The basic repetitive cell consists of three reflective ribbons [see Figure 5.19(b)]: one moveable ribbon, a reference "ribbon," and a compensating "ribbon," with the latter two being spatially fixed and separated by an integer multiple of half the center wavelength ($N\lambda_0/2$) where λ_0 is typically around 1,550 nm (i.e., their phasors will be in phase only at the center wavelength). In the nominal undeflected state, all three phasors have the same orientations at the center wavelength λ_0 and add constructively to reflect the light without diffraction (no attenuation by the VOA). Pulling the moveable ribbon down by $\lambda_0/4$ adds a round trip phase of π at the center



(a)



(b)

Figure 5.19 (a) Phasor description of the diffractive operation of the variable optical attenuator. At the center wavelength, the phasors add to the null vector. At other wavelengths, the compensating ribbon introduces an error vector that cancels the error vector introduced by the moveable ribbon, thus providing broadband achromatic operation [42]. (b) A schematic illustration of the achromatic implementation of the variable optical attenuator. The structure consists of groups of three ribbons, one of which is moveable and two of which are spatially fixed. The latter two are vertically separated by $N\lambda_0/2$ where λ_0 is the center wavelength and N is an integer.

wavelength to the light reflected by this ribbon. Schematically, the corresponding phasor, ε_m, rotates in the complex plane by 180°. At the center wavelength, ε_c, the phasor corresponding to the compensating ribbon remains in the same orientation as ε_r, the phasor for the reference ribbon. The three phasors now add destructively to a null vector [see Figure 5.19(a)] at the center wavelength, and thus light diffracts into higher orders, causing maximum attenuation of the main undiffracted order. At a wavelength λ different than λ_0, the phasor ε_m rotates by an amount $\pi\lambda_0/\lambda$ radians (less or more than π), causing an error vector relative to the phasor at λ_0. Simultaneously, the phasor ε_c rotates by $2N\pi\lambda_0/\lambda$, causing an error vector in the opposite direction—ε_c rotates past ε_m by an additional $\pi\lambda_0/\lambda$ (if $N = 1$), placing it in an opposite quadrant to ε_m. As the magnitudes of the phasors are proportional to the areas of the ribbons, the two error vectors can be made to cancel each other out under certain geometrical conditions. Analytical calculations show that if A_m, A_r, and A_c are the respective areas of the moveable, reference and compensating ribbons, then there are two conditions that must be satisfied: $A_r+2A_c = A_m$ and $A_c/A_m = 1/2N$. The first condition ensures equality of the magnitudes of the phasors that are out of phase. The second condition follows from matching the phases of the error vectors. As a result, the total phasor is null ($\varepsilon_m + \varepsilon_{mr}+\varepsilon_c = 0$) over a wide range of wavelengths.

Extending the achromatic design to also eliminate polarization dependence entails mapping the linear geometry (linear ribbons) into one with cylindrical symmetry (circular discs), making the device effectively a two-dimensional phase grating (see Figure 5.20). The reference ribbon becomes a reference circular post; the moveable ribbon becomes a membrane with circular cut outs suspended by anchor points on the edges; and the achromatic compensating ribbons become annular rings around the reference posts. The membrane incorporates minute release holes that assist in the fast and uniform removal of the sacrificial layer during fabrication. The dimensions of the gaps remain unchanged.

In a typical design, N equals 3, the center wavelength is 1,550 nm, corresponding to a height difference between the moveable membrane and compensating annuli of 2.32 μm. The periodicity of the repeating diffractive element is typically between 20 and 200 μm [42]. The widths of the reference post, as well as the gap between the post and membrane, are typically a few micrometers. The resulting variable optical

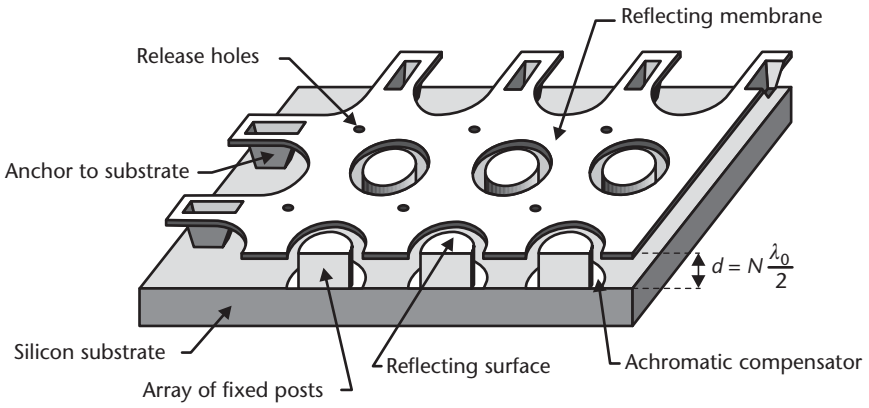


Figure 5.20 A cross-sectional schematic of the variable optical attenuator. The architecture incorporates achromatic compensation and cylindrical symmetry to ensure low dependence on polarization [42].

attenuator from Lightconnect has a dynamic range (attenuation range) of 30 dB, a wavelength dependence of attenuation of 0.25 dB, and a polarization-dependent loss of 0.2 dB. The total insertion loss, which includes losses from fiber coupling, is 0.7 dB. The response time of the device is, as expected from the GLV, quite fast, measuring 40 μs. The actuation voltage between the membrane and substrate is less than 8V. The company also provides a specification for reliability: in excess of 100 billion cycles for wear out. While wear out is very subjective and not quantified, it reflects the projected reliability of this device where displacements are very small ($\lambda_0/4 \approx 400$ nm) and friction is nonexistent.

The fabrication is very similar to that of the GLV with a few exceptions. First, lithography followed by an etch defines the reference posts with a height of 2.32 μm. A thin (20–60 nm) layer of silicon dioxide is thermally grown. A layer of sacrificial polysilicon or amorphous silicon is deposited. This layer must be optically smooth, as any defects will subsequently imprint the moveable membrane. Holes are etched through the sacrificial layer to allow for the anchor points to the substrate. Silicon nitride is then deposited as the membrane material. It may be stochiometric or silicon rich. A lithographic step followed by an etch step pattern the nitride layer into the desired membrane layout. Finally, xenon difluoride (XeF_2) removes the sacrificial layer of silicon to release the membrane. A subsequent evaporation step deposits a thin gold layer across the entire surface, ensuring high reflectivity in the infrared.

Summary

This chapter reviewed a number of commercially available products with applications in imaging, displays, and fiber-optical telecommunications. The applications are very diverse but share the common use of MEMS technology to manipulate light. While MEMS have proven to be vital for the operation of the aforementioned products, it remains an enabling technology and a means to an end. It is imperative to understand the final application in order to assess the importance and applicability of MEMS for that particular application.

References

[1] Cole, B. E., R. E. Higashi, and R. A. Wood, "Monolithic Two-Dimensional Arrays of Micromachined Microstructures for Infrared Applications," in *Integrated Sensors, Microactuators, & Microsystems* (*MEMS*), K. D. Wise (ed.), *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1679–1686.

[2] Van Kessel, P. F., et al., "A MEMS-Based Projection Display," in *Integrated Sensors, Microactuators, & Microsystems* (*MEMS*), K. D. Wise (ed.), *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1687–1704.

[3] Bloom, D. M., "The Grating Light Valve: Revolutionizing Display Technology," *Proc. SPIE, Projection Displays III*, Vol. 3013, San Jose, CA, February 10–12, 1997, pp. 165–171.

[4] Chen, Y., et al., "Metro Optical Networking," *Bell Labs Technical Journal*, January–March 1999, pp. 163–186.

[5] Tomsu, P., and C. Schmutzer, *Next Generation Optical Networks*, Upper Saddle River, NJ: Prentice Hall, 2002, pp. 68–70.

[6] Saleh, B. E. A., and M. C. Teich, *Fundamentals of Photonics*, New York: Wiley, 1991, pp. 461–466, 494–503.

[7] Siegman, A. E., *Lasers*, Mill Valley, CA: University Science Books, 1986, pp. 1–80.

[8] Coldren, L. A., and S. W. Corzine, *Diode Lasers and Photonic Integrated Circuits*, New York: Wiley, 1995, pp. 1–9, 111–116.

[9] Saleh, B. E. A., and M. C. Teich, *Fundamentals of Photonics*, New York: Wiley, 1991, pp. 310–317.

[10] Jerman, H., and J. D. Grade, "A Mechanically-Balanced, DRIE Rotary Actuator for a High-Power Tunable Laser," *Tech. Digest Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, June 2–6, 2002.

[11] Pezeshki, B., et al., "20 mW Widely Tunable Laser Module Using DFB Array and MEMs Selection," *IEEE Photonics Technology Letters*, Vol. 14, October 2002, pp. 1457–1459.

[12] Littman, M. G., and H. J. Metcalf, "Spectrally Narrow Pulsed Dye Laser Without Beam Expander," *Applied Optics*, Vol. 17, No. 14, 1978, pp. 2224–2227.

[13] Agrawal, G. P., and N. K. Dutta, *Semiconductor Lasers*, Boston, MA: Kluwer Academic Publishers, 1993, pp. 269–275.

[14] Coldren, L. A., and S. W. Corzine, *Diode Lasers and Photonic Integrated Circuits*, New York: Wiley, 1995, pp. 17–24, 393–398.

[15] Klein, M. V., *Optics*, New York: Wiley, 1970, pp. 342–346.

[16] Klein, M. V., *Optics*, New York: Wiley, 1970, pp. 338–341.

[17] Liu, K., and M. G. Littman, "Novel Geometry for Single-Mode Scanning of Tunable Lasers," *Optics Letters*, Vol. 6, No. 3, 1981, pp. 117–118.

[18] U.S. Patent 6,469,415, October 22, 2002.

[19] Smith, S. T., and D. G. Chetwynd, *Foundations of Ultraprecision Mechanism Design (Developments in Nanotechnology)*, London, UK: Taylor and Francis, 1992, p. 119.

[20] Tang, W. C., et al., "Electrostatic-Comb Drive of Lateral Polysilicon Resonators," *Sensors and Actuators*, Vol. A21, Nos. 1–3, February 1990, pp. 328–331.

[21] Berger, J. D., and D. Anthon, "Tunable MEMS Devices for Optical Networks," *Optics & Photonics News*, March 2003, pp. 43–49.

[22] Kogelnik, H., and C. V. Shank, "Coupled-Wave Theory of Distributed Feedback Lasers," *Journal of Applied Physics*, Vol. 43, 1972, pp. 2327–2335.

[23] Data sheet for CQF935/508 series, JDS Uniphase Corporation, 1768 Automation Parkway, San Jose, CA 95131, http://www.jdsu.com.

[24] Ghafouri-Shiraz, H., *Distributed Feedback Laser Diodes and Optical Tunable Filters*, New York: Wiley, 2003.

[25] Amann, M. -C., and J. Buus, *Tunable Laser Diodes*, Norwood, MA: Artech House, 1998, pp. 40–51.

[26] Pezeshki, B., et al., "Twelve Element Multi-Wavelength DFB Arrays for Widely Tunable Laser Modules," *Tech. Digest of the Optical Fiber Communication Conference*, Anaheim, CA, March 17–22, 2002, pp. 711–712.

[27] Saleh, B. E. A., and M. C. Teich, *Fundamentals of Photonics*, New York: Wiley, 1991, pp. 316–317.

[28] Plomteux, O., "DFL-5720 Digital Frequency-Locking System: Simplifying Wavelength-Locker Testing," Application Note 083, EXFO Electro-Optical Engineering, Inc., Vanier, Quebec, Canada, http://documents.exfo.com/appnotes/anote083-ang.pdf.

[29] Dames, M. P., et al., "Efficient Optical Elements to Generate Intensity Weighted Spot Arrays: Design and Fabrication," *Applied Optics*, Vol. 30, No. 19, July 1, 1991, pp. 2685–2691.

[30] Farn, M. W., "Agile Beam Steering Using Phase-Array Like Binary Optics," *Applied Optics*, Vol. 33, No. 22, August 1, 1994, pp. 5151–5158.

[31] Hecht, J., *Understanding Fiber Optics*, 3rd ed., Upper Saddle River, NJ: Prentice Hall, 1999, pp. 133–134, 320–325, 373–374, 455.

[32] Marxer, C., et al., "Vertical Mirrors Fabricated by Deep Reactive Ion Etching for Fiber-Optic Switching Applications," *Journal of Microelectromechanical Systems*, Vol. 6, No. 3, September 1997, pp. 185–277.

[33] Hecht, J., *Understanding Fiber Optics*, 3rd ed., Upper Saddle River, NJ: Prentice Hall, 1999, pp. 62–72.

[34] Zou, J., et al., "Optical Properties of Surface-Micromachined Mirrors with Etch Holes," *Journal of Microelectromechanical Systems*, Vol. 8, No. 4, December 1999, pp. 506–513.

[35] Iannone, E., and R. Sabella, "Optical Path Technologies: A Comparison Among Different Cross-Connect Architectures," *Journal of Lightwave Technology*, Vol. 14, No. 10, October 1996, pp. 2184–2196.

[36] U.S. Patents 5,629,790, May 13, 1997; 6,480,320 B2, November 12, 2002; and 6,628,041 B2, September 30, 2003.

[37] Burns, B., et al., "Electromagnetically Driven Integrated 3D MEMS Mirrors for Large Scale PXCs," in *Proceedings of National Fiber Optics Engineers Conference*, NFOEC 2002, Dallas, TX, September 15–19, 2002.

[38] Saleh, B. E. A., and M. C. Teich, *Fundamentals of Photonics*, New York: Wiley, 1991, pp. 81–105.

[39] Temesvary, V., et al., "Design, Fabrication, and Testing of Silicon Microgimbals for Super-Compact Rigid Disk Drives," *Journal of Microelectromechanical Systems*, Vol. 4, No. 1, March 1995, pp. 18–27.

[40] Hecht, J., *Understanding Fiber Optics*, 3rd ed., Upper Saddle River, NJ: Prentice Hall, 1999, pp. 99–100.

[41] Agrawal, G., *Nonlinear Fiber Optics*, 2nd ed., San Diego, CA: Academic Press, 1995, pp. 239–243, 316–399.

[42] U.S. Patents 6,169,624, January 2, 2001, and 6,501,600, December 31, 2002.

[43] Halliday, D., and R. Resnick, *Physics*, 3rd ed. extended, New York: Wiley, 1988, pp. 907–910.

Selected Bibliography

Buser, P., and M. Imbert (translated by R. H. Kay), *Vision*, Cambridge, MA: The MIT Press, 1992.

Hecht, J., *Understanding Fiber Optics*, 3rd ed., Upper Saddle River, NJ: Prentice Hall, 1999.

MacDonald, L. W., and A. C. Lowe (eds.), *Display Systems: Design and Applications*, West Sussex, England: Wiley, 1997.

Micromechanics and MEMS: Classic and Seminal Papers to 1990, W. Trimmer (ed.), New York: IEEE, 1997.

Wise, K. D. (ed.), "Special Issue on Integrated Sensors, Microactuators, and Microsystems (MEMS)," *Proceeding of the IEEE*, Vol. 86, No.8, August 1998.

CHAPTER 6

MEMS Applications in Life Sciences

> "Jim, you've got to let me go in there! Don't leave him in the hands of Twentieth-Century medicine."
>
> —*Dr. Leonard McCoy speaking to Captain James Kirk, in the movie* Star Trek IV: The Voyage Home*, 1986.*

The "medical tricorder" in the famed *Star Trek* television series is a purely fictional device for the remote scanning of biological functions in living organisms. The device remains futuristic, but significant advances in biochemistry have made it possible to decipher the genetic code of living organisms. Today, dozens of companies are involved in biochemical analysis at the microscale, with a concentration of them involved in genomics, proteomics, and pharmacogenics. Their successes have already had a positive impact on the health of the population; examples include faster analysis of pathogens responsible for illness and of agricultural products as well as more rapid sequencing of the human genome. Systems expected in the near future will detect airborne pathogens responsible for illness (such as Legionnaire's disease or anthrax in a terrorist attack) with a portable unit, give on-demand genetic diagnostics for the selection of drug therapies, be able to test for food pathogens such as *E. coli* on site, and more rapidly test for bloodborne pathogens.

Conventional commercial instruments for biochemical and genetic analysis, such as those available from Applied Biosystems of Foster City, California, perform a broad range of analytical functions but are generally bulky. The concept of *micro total analysis system* (μTAS), which aims to miniaturize all aspects of biochemical analysis, with its commensurate benefits, was introduced in 1989 by Manz [1]. This chapter begins with an introduction to microfluidics, followed by descriptions of the state of the art of some of the microscale methods used in DNA analysis. Finally, electrical probe techniques and some applications are presented. A common theme will be the use of glass and plastic substrates, in contrast to most of the devices in other chapters of this book.

Microfluidics for Biological Applications

The biological applications of MEMS (bio-MEMS) and microfluidics are inextricably linked because the majority of devices in systems for biological and medical analysis work with samples in liquid form. Outside of biological analysis, microfluidics have applications in chemical analysis, drug synthesis, drug delivery, and point-of-use synthesis of hazardous chemicals. In this section, we discuss common pumping methods in bio-MEMS and the issue of mixing.

Pumping in Microfluidic Systems

Examples of flow channels used in microfluidics are rectangular trenches in a substrate with cap covers on top, capillaries, and slabs of gel, having cross-sectional dimensions on the order of 10 to 100 μm and lengths of tens of micrometers to several centimeters. For microfluidic biological analysis, fluid drive or pumping methods include applied pressure drop, capillary pressure, electrophoresis, electroosmosis, electrohydrodynamic force, and magnetohydrodynamic force; the first four are common. Pressure drive, the most familiar from the macroscopic world, is simply the application of a positive pressure to one end of a flow channel. Alternatively, a negative pressure (vacuum) can be applied to the other end. Due to drag at the walls, the flow is slowest at the edges, increasing in a parabolic profile to a maximum at the center [see Figure 6.1(a)].

Another familiar pumping force is the wicking action of small-diameter capillaries. This force is due to surface tension (i.e., the surface energy of the system can be lowered if the solid-gas interface is replaced by a solid-liquid interface). Capillary action is commonly used to load liquid into a channel. After insertion of the end of a

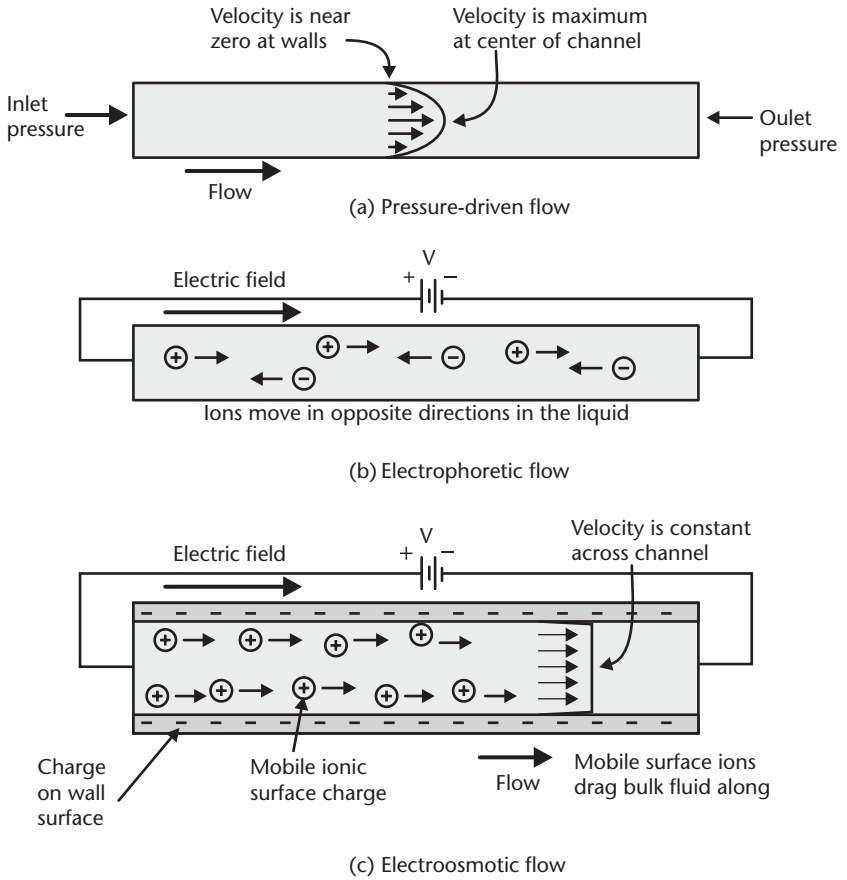


Figure 6.1 Three types of pumping used in microfluidics: (a) pressure drive, in which a pressure forces the volume fluid to flow; (b) electrophoretic flow, in which ions of opposite polarity in solution flow in opposite directions under the effect of externally applied electric field; and (c) electroosmotic flow, in which an electric field moves the mobile ion sheath of the surface double layer, dragging the volume in the channel along with it.

capillary into a larger container of sample, or addition of liquid to a well at the mouth of a channel on a chip, liquid is drawn into the channel without the application of additional pressure.

Electrophoretic flow can be induced only in liquids or gels with ionized particles. The application of a voltage across the ends of the channel produces an electric field along the channel that drives positive ions *through* the liquid toward the negative terminal and the negative ions to the positive terminal [see Figure 6.1(b)]. Neutral particles in the channel are not directly affected by the field. The velocity of the ions is proportional to the electric field and charge and inversely related to their size [2]. In liquids, velocity is also inversely related to the viscosity, while in gels the velocity depends on porosity.

Electroosmotic flow occurs because channels in glasses and plastics tend to have a fixed charge on their surfaces. In glasses, silanol (SiOH) groups at the walls are deprotonated in solution (they lose the hydrogen as a positive ion), leaving the surface with a negative charge [3]. These negative ions then attract a diffuse layer of positive ions, forming a *double layer* in the liquid [see Figure 6.1(c)]. The layer of positive ions is not tightly bound and can move under an applied electric field. When this sheath of ions moves, it drags the rest of the channel volume along with it, creating electroosmotic flow. In contrast to pressure-driven flow, the velocity at the center of the channel is about the same or slightly less, giving the fluid a flat velocity profile. This *plug flow* is advantageous in many situations in biological analysis where the spreading of a short-length sample into neighboring regions of a channel is not desired. Electroosmotic pumping works best with small-dimension channels. Flow velocities can range from a few micrometers per second to many millimeters per second.

Electrophoretic flow and electroosmotic flow can be grouped together under the heading of electrokinetic flow; indeed, both occur simultaneously in ionic solutions with an applied electric field. The one that dominates depends on the details of the solution and walls. Manufacturers of analysis equipment employing electrokinetic flow generally design the system so that only one dominates. For example in gel electrophoresis, the solution is a porous gelatinous medium, which cannot move as a liquid would in electroosmosis. Instead, the charges percolate electrophoretically under the effect of the electric field through the porous gel. Alternatively, a liquid buffer solution can be used in microchannels. Electroosmosis can dominate, pushing the bulk of the flow in one direction. Positive ions within this bulk flow move even faster relative to the bulk solution, while negative ions move in the opposite direction with respect to the bulk solution, giving them a slower net velocity [3].

Mixing in Microfluidics

Volumetric flow rates in microscale channels are of course much lower than in macroscopic channels, such as the water pipes in a building. The *Reynolds number* is useful for comparing flows of different fluids in channels of dimensions that vary over orders of magnitude. The Reynolds number is a dimensionless number related to the ratio of kinetic energy in the fluid to the rate of loss of energy to friction. It is given by $\rho \bullet v \bullet D/\mu$, where ρ is the fluid density, v is the average velocity, D is the diameter or equivalent "hydraulic diameter" of the channel, and ρ is the absolute viscosity. For Reynolds numbers below about 2,300 for a tube with circular cross

section, flow is laminar: the fluid can be envisioned as flowing in laminar sheets, moving slowest at the edges due to the drag of the walls and moving fastest at the center. For higher Reynolds numbers, the flow is turbulent rather than laminar. In microfluidics, water-based solutions are usually used, having $\rho \approx 1$ g/cm^3 and $\mu \approx 0.01$ g/(cm•s). For a representative hydraulic diameter of 30 μm and a representative velocity of 1 mm/s, the Reynolds number is merely 0.03. In microfluidics, Reynolds numbers are usually below one [4].

This has great implications for mixing in microfluidics. In the macroscopic world, simply joining two channels together would enable the two streams to intermix. At these low Reynolds numbers, however, streams joined from two channels simply flow side by side, with intermixing only by diffusion. This is used to advantage in the Agilent Cell LabChip®, which detects cells stained with fluorescent dyes. When placed in the Agilent 2100 Bioanalyzer system, a vacuum pulls separate flows of cells and buffer together in a Y-shaped junction (see Figure 6.2). The flow of cells is pushed to one side of the microchannel by the flow of buffer. Individual stained cells are detected as they pass under an excitation beam and fluoresce. This concentration scheme is used because individual cells would clog a flow channel of the same width. Often the opposite situation, mixing, is desired. In this case, special flow structures, which add some turbulence or increase the area of diffusive mixing, have been demonstrated to overcome this problem [5].

DNA Analysis

The Structure of DNA

The genetic code is stored in cell chromosomes, each containing long strands of *deoxyribonucleic acid* (DNA) [6, 7]. The building blocks of DNA are molecules called *nucleotides* that consist of a "base" joined to a sugar-phosphate backbone [see Figure 6.3(a)]. The nomenclature often interchanges between base and nucleotide to represent the same building block. In DNA there are four types of nucleotides differentiated by their bases: *adenine*, *thymine*, *cytosine*, and *guanine*. The nucleotides are labeled according to the first letter of their corresponding bases: A, T, C,

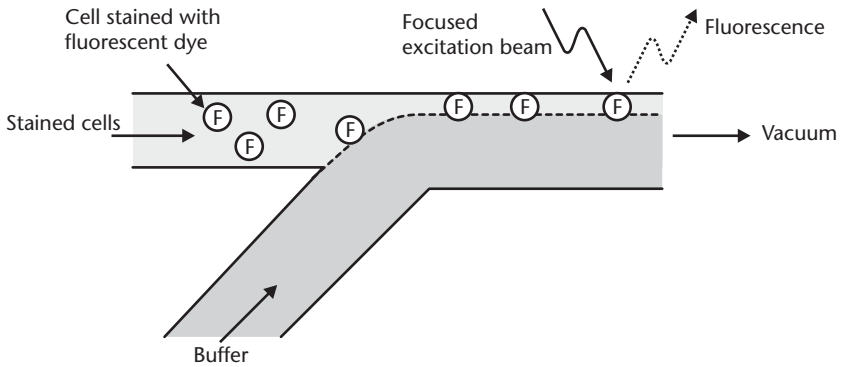


Figure 6.2 Example of the use of laminar flow in microfluidics: In the Cell LabChip from Agilent Technologies of Palo Alto, California, the flow of cells tagged with a fluorescent dye is pushed to one side of the channel. Individual cells are detected when they fluoresce.

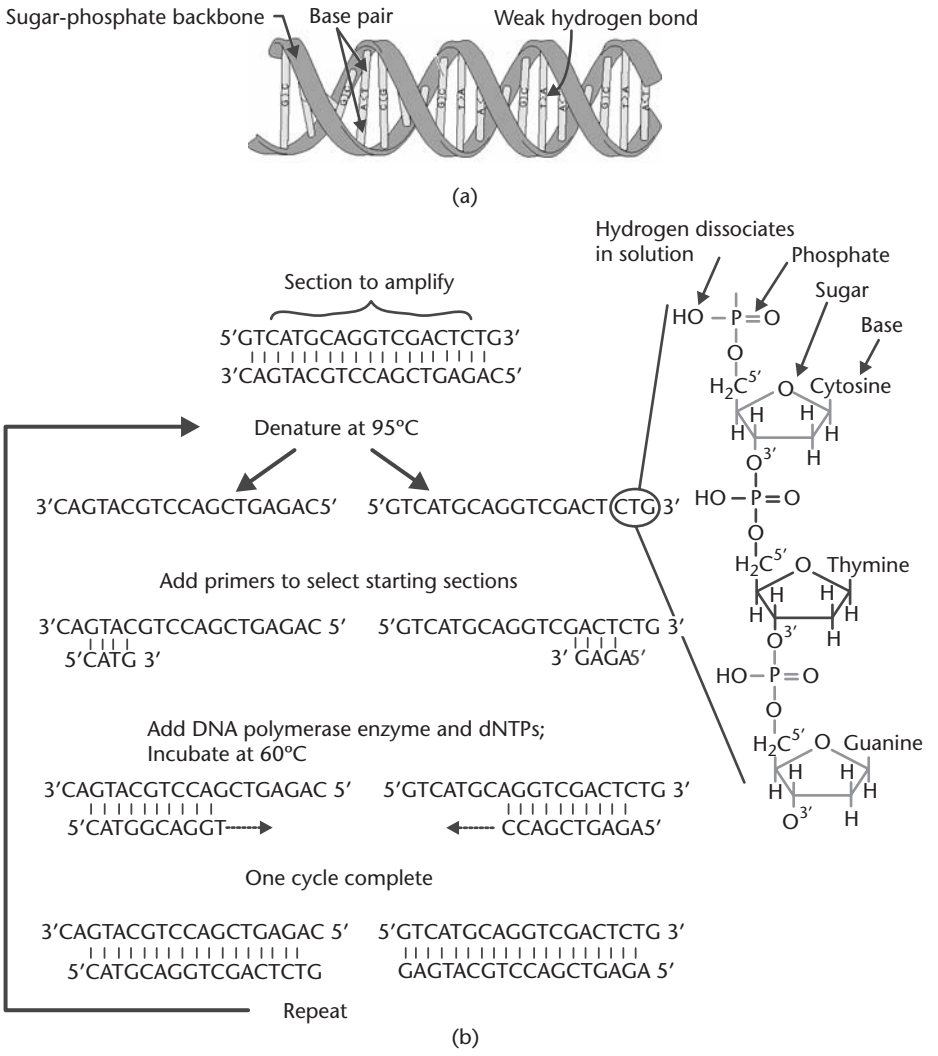


Figure 6.3 Illustration of (a) the twisted double-helix structure of DNA; and (b) the polymerase chain reaction (PCR). Denaturing of the starting DNA template at 95°C yields two strands, each containing all of the necessary information to form a complementary replica. The addition of primers defines the starting point for replication. At 60°C, the DNA polymerase enzyme catalyzes the reconstruction of the complementary DNA strand from an ample supply of nucleotides (dNTPs). The reconstruction always proceeds in the 5′→3′ direction. The cycle ends with copies of two portions of the helices, in addition to the starting template. The cycle is then repeated. The exploded view of three nucleotides (CTG) in the denatured template shows their chemical composition, including the 3′-hydroxyl and 5′-phosphate groups. (*After:* [6, 7].)

and G, respectively. This is the four-letter alphabet of DNA. The human genome has 23 separate pairs of chromosomes, averaging 130 million base pairs in length, for a total of about three billion base pairs. Genes that form the template for proteins are typically 27,000 base pairs long, but only about 1,000 are used; the rest are extra "filler" bases.

Each nucleotide molecule has two ends, labeled 3' and 5', corresponding to the hydroxyl and phosphate groups attached to the 3' and 5' positions of carbon atoms in the backbone sugar molecule [see Figure 6.3(b)]. In the long DNA chain, the 3'

end of one nucleotide connects to the 5' end of the next nucleotide. This essentially gives directionality to the DNA chain.

Two strands of DNA are joined by weak hydrogen bonds to form the well-known twisted double-helix structure [6]. The attachment occurs between specific pairs of nucleotides: guanine bonds to cytosine (G–C), and adenine bonds to thymine (A–T). This important pairing property is known as *complementarity*. Color photography makes a simple analogy to understand complementarity: The three additive primary colors—red, green, and blue—are in their respective order complementary to the three subtractive colors—cyan, magenta, and yellow. A positive photographic print and its negative contain the same image information, even though the colors of the positive (the additive colors) are different from the colors of the negative (the subtractive colors). The positive and negative in photography are analogous to the two complementary strands of DNA in a double helix.

PCR

A primary objective of genetic diagnostics is to decipher the sequence of nucleotides in a DNA fragment after its extraction and purification from a cell nucleus. This task is difficult due to the miniscule concentration of DNA available from a single cell. As a solution, scientists resort to a special biochemical process called *amplification* to create a large number of identical copies of a single DNA fragment. The most common amplification method is the polymerase chain reaction (PCR). Invented in the 1980s by Kary Mullis, for which he was awarded the Nobel Prize in Chemistry in 1993, it allows the replication of a single DNA fragment using complementarity. The basic idea is to physically separate—denature—the two strands of a double helix and then use each strand as a template to create a complementary replica.

The polymerase chain reaction begins by raising the temperature of the DNA fragment to 95°C in order to denature the two strands. Incubation occurs next at 60°C in a solution mix containing a special enzyme (called DNA polymerase, an example of which is Taq polymerase), an ample supply of nucleotides (dNTPs), and two complementary primers. The primers are short chains of nucleotides previously synthesized to hybridize—or to specifically match up using complementarity—with a very small segment of the longer DNA fragment and consequently define the starting point for the replication process. The DNA polymerase enzyme catalyzes the construction of the complementary DNA strand beginning from the position of the primer and always proceeding in the 5' → 3' direction. Replication of a portion of the single strand is rapid, proceeding at a rate of about 50 bases per second [8]. The cycle ends with two identical copies of only the sections between (and including) the primers, in addition to the starting DNA template. Repetition of the cycle increases the number of identical copies with a factor of 2^n, where n is the number of cycles; thus, after 20 cycles, about one million copies have been created. The efficiency drops after about 20 cycles [9], but 30 to 40 cycles are typically needed to generate sufficient product for later analysis.

PCR on a Chip

There are several advantages to miniaturizing the PCR process. Smaller chambers have a greater ratio of surface area to volume. Surface area affects the rate of heat

conduction, and volume determines the amount of heat necessary for a thermal cycle. A greater ratio of surface area to volume, therefore, enables faster thermal cycling in PCR. Because the chamber volume is smaller, less sample and volume of expensive reagents is needed. If integrated with a detection scheme such as electrophoretic separation or TaqMan® tagging (described later) on the same chip, the entire process is simplified, making it faster, less expensive, and more repeatable.

PCR on a silicon chip was first demonstrated around 1994 by several groups [10, 11], and by the end of the 1990s there had been several demonstrations of PCR on a chip. This section describes silicon miniature PCR thermal cycling chambers developed at Lawrence Livermore National Laboratory (LLNL) of Livermore, California (see Figure 6.4) [12]. Different versions of this chamber are at the core of portable analytical instruments under development at Cepheid of Sunnyvale, California, and Microfluidic Systems, Inc., of Pleasanton, California.

Several generations of micromachined chambers have been fabricated at LLNL [13]. They thermally cycle a solution between the denaturing and incubation temperatures, approximately 95°C and 60°C, respectively. One chamber, with a volume of 25 to 100 μl, is made of two silicon chips with etched grooves, which are bonded together. A silicon nitride window provides optical access. Experimental results have shown that bare silicon inhibits PCR amplification, so a disposable polypropylene liner was added to the chamber. This slows the rate at which the chamber can be heated and cooled slightly from an all-silicon version to about 8°C/s. An advantage of a disposable liner is that the chamber no longer has to be cleaned. Eliminating this time-consuming operation enables more samples to be run per day.

Earlier designs had a polysilicon heater on a silicon nitride membrane for heating the fluid inside the chamber and used a separate, external temperature sensor. By changing the heater material to platinum, which is commonly used as a temperature sensor, both heating and sensing operations can be performed with the same platinum element. Testing of early devices showed that there were temperature variations as high as 10°C across the chamber. By relocating the heater away from

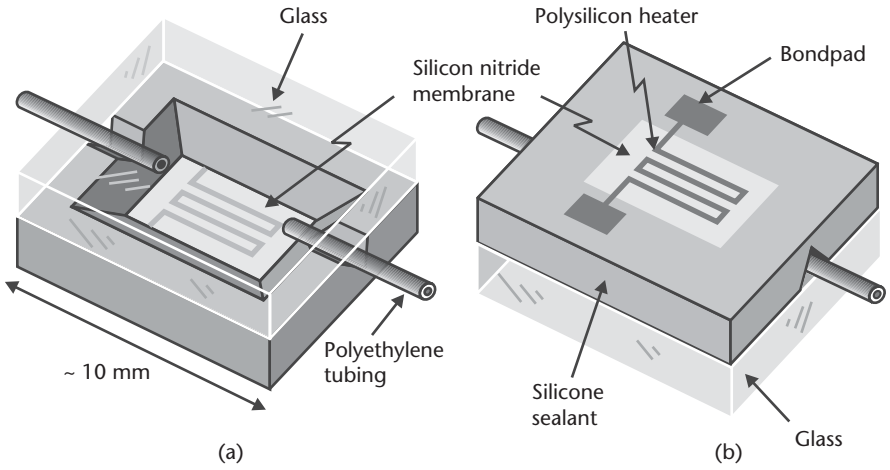


Figure 6.4 Illustrations of (a) the front side, and (b) the back side of an early micromachined silicon PCR chamber. A polysilicon heater on a silicon nitride membrane cycles the solution between the denaturing and incubation temperatures of PCR. (*After:* [12].)

the membrane so that heat flows through the highly thermally conductive silicon walls of the chamber, the temperature uniformity of the fluid was greatly improved. A fan was added for more rapid cooling. These modifications have yielded much tighter closed-loop temperature control and enabled faster cycling, from around 35s per cycle to as little as 17s per cycle. These cycle times are far faster than the approximately 4 min per cycle needed in the industry-workhorse Applied Biosystems GeneAmp® PCR System 9600 [13].

The LLNL system has detection capability in addition to amplification. In a variation of traditional PCR, the addition of TaqMan dyes (probes), which link to certain sections of a DNA strand (just like the primers), results in fluorescence of green light from each replicated DNA strand when excited by a blue or ultraviolet source [13, 14]. Thus, the intensity of the fluorescence is proportional to the number of replicated DNA strands matching the TaqMan probe in the solution. This procedure has the advantage of simultaneous DNA amplification and detection but only works when suitable primers and probe have been added to the solution for the type of DNA under test. Thus, the number of different DNA sections potentially being identified is equal to the number of PCR chambers that can be run simultaneously. In demonstrations at LLNL with different cells, there was no detectable fluorescence signal for the first 20–25 cycles, depending on the initial concentration. After cycling on the order of 5–15 minutes, the signal appeared and rapidly grew if there was a match.

In the LLNL system, the light source is a filtered blue LED through the silicon nitride window. A handheld prototype, which represents the holy grail of DNA analysis, is about the size of a one-quart milk carton, including computer, display, and keypad, and is powered by a separate 0.5-kg battery with a run time of two hours. Larger but still portable systems using this technology, available from Microfluidic Systems, can presently identify over 10 airborne pathogens.

Electrophoresis on a Chip

Determining the sequence of nucleotides in a DNA strand involves amplification and chemical labeling of the amplified DNA fragments with specific fluorescent or radioactive tags and a subsequent distinct detection step that analyzes the labeled DNA products. The entire process is called *DNA sequencing*. Its underlying principles are beyond the scope of this book, but the eager reader is referred to Stryer's book on biochemistry [6]. One detection technique is electrophoresis, which employs the separation of charged molecules, including DNA, in suspension under the effect of an electric field [see Figure 6.5(a)]. In solution, a hydrogen ion dissociates from each phosphate in the DNA backbone, leaving the DNA strand with a net negative charge [see Figure 6.3(b)]. The charge-to-mass ratio is approximately the same for strands of different lengths, but, when driven with an electric field through a molecular sieve, larger molecules move more slowly [9]. Thus, after a given time, groups of small molecules move farther than larger ones. A limitation of electrophoresis is that as the sample sits in solution, it is also diffuses both up and down the channel. Because the diffusion distance grows with time, short electrophoretic separation and detection times are advantageous, which implies the use of a high electric field over a short distance. Electrophoresis can separate DNA fragments up to about 3,000 bases in length.

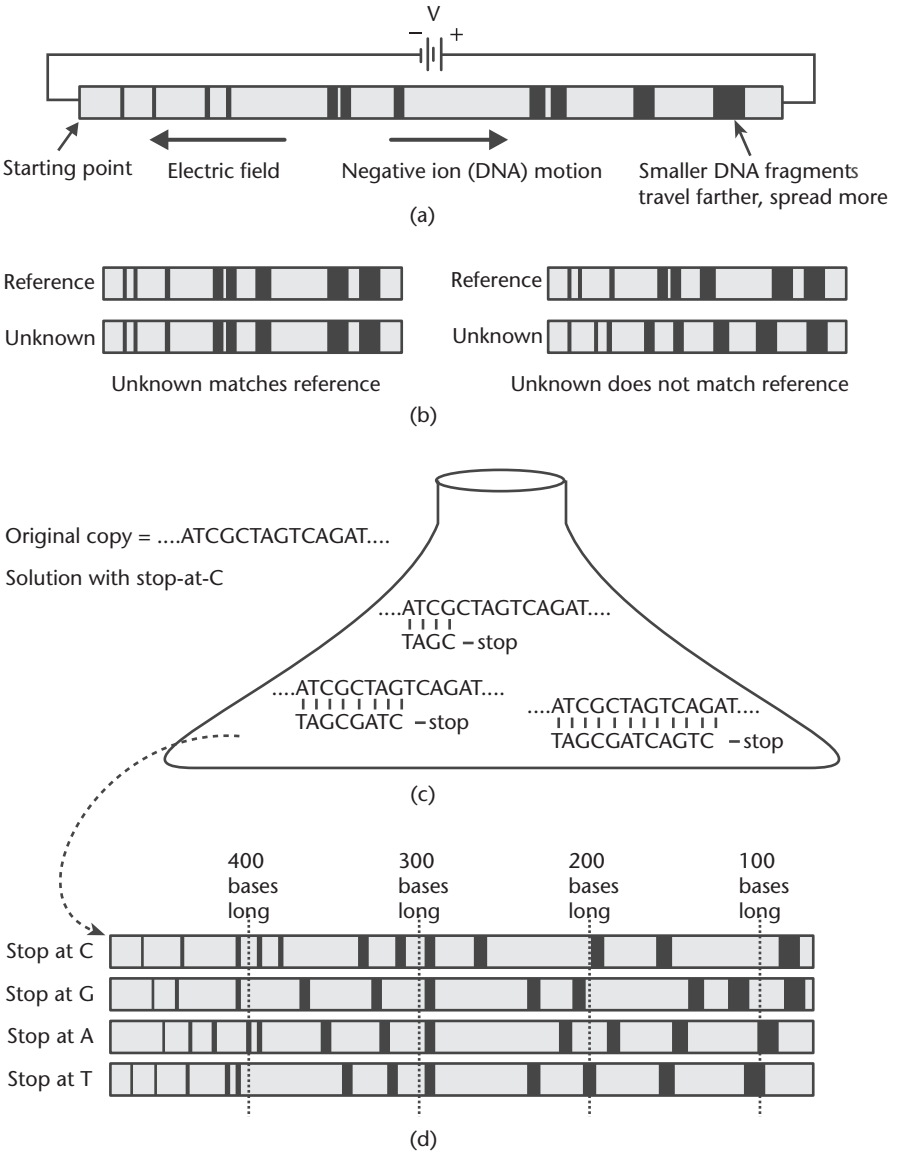


Figure 6.5 (a) Illustration of electrophoresis to sort DNA fragments by size. Here, a sieving medium is assumed so that negative charges move to the right. Charged molecules move under the effect of the applied electric field. (b) Comparison of known and unknown samples based on fragment length. (c) Illustration of the Sanger method: copies are made of the original DNA, randomly stopping at the same nucleotide (C in this example) to produce variable-length fragments with the same ending. (d) Fragments with each ending undergo electrophoresis.

In *gel electrophoresis*, DNA products are introduced at the edge of a porous gelatinous sheet that is 20 to 100 cm long. The electric field is limited to only 5–40 V/cm due to Joule heating [9]. In *capillary electrophoresis* [15], the products are fed into a thin capillary tube, 10 to 300 μm in diameter and approximately 50 cm long, with an applied electric field of up to 1,200 V/cm [9]. Higher fields can be used with smaller cross sections due to the ability to remove heat more rapidly. Before electrophoresis is performed, the DNA strands are processed to add a tag for later

detection. One type of tag is radioactive (^{32}P), which is imaged with photographic film to determine the position of the strand in the gel or capillary. A more common tag added to the 5' end fluoresces under ultraviolet excitation, emitting light at a visible wavelength. Used alone, electrophoretic separation can compare two samples of fragments of DNA to determine whether they match but cannot tell the exact sequence.

If electrophoresis is to be employed to determine the sequence of bases on a section of DNA, the Sanger method may be used for fragments up to about 1,000 bases long [see Figure 6.5(c, d)] [9]. This begins with many identical copies of single, denatured sections of DNA. Replication in a solution with dNTPs is started from the 5' end, just as in PCR. In this case, however, a small concentration of bases in the solution of one type, such as C (cytosine), is altered so that the replication of that DNA strand stops when the replication-halting base is used. This results in copies of the original strands of varying length that always end in C. The same is done in separate solutions with small concentrations of replication-halting bases of the other types (G, A, and T). The four groups of variable-length copies then undergo electrophoresis in four parallel channels. Sequences of each length, from one base to the maximum in the original sample, are separated for reading, and the results from the four channels are compared to infer the entire sequence of the strand.

Miniaturization brings many benefits to capillary electrophoresis. The length of the sample emitted into the channel can be kept relatively short (on the order of 100 μm), reducing the distance that must be traveled for the fragments of different lengths to separate. Reducing the length of the channel decreases the applied voltage required to maintain a high electric field from a few kilovolts down to hundreds of volts. Faster separation times also become possible because the molecules have to travel shorter distances. Additionally, the overall volume of DNA and reagents decreases significantly to one microliter or less.

Early demonstrations of capillary electrophoresis on a chip took place in 1992 at Ciba-Geigy, Ltd., of Basle, Switzerland [16]. Woolley and Mathies [17, 18] from the University of California, Berkeley, were the first in 1994 to demonstrate DNA sequencing by capillary electrophoresis on a glass chip. The structure of their device consists of two orthogonal channels etched with buffered hydrofluoric acid into a first glass substrate: a short channel for injecting fluid and a long channel for separating the DNA fragments (see Figure 6.6). A second glass substrate covers the channels and is secured to the first substrate with an intermediate adhesive or by thermal bonding. Holes etched or drilled with a diamond-core drill in the top glass substrate provide fluid access ports to the embedded channels. Both channels are typically 50 μm wide and 8 μm deep but can be as wide as 100 μm and as deep as 16 μm; the separation channel is 3.5 cm long. Thermal bonding is achieved by ramping the temperature of the glass plates in an oven to 600°C at the rate of 5°C/min, holding the temperature for 2 to 3 hours, then ramping down to room temperature [18]. The surfaces of the channels have a coating to eliminate charging due to deprotonation, preventing electroosmosis from occurring. The injection and separation channels are filled with sieving matrix of hydroxyethylcellulose by applying a vacuum to one end.

The fluid containing the DNA fragments is admitted into the injection channel, and the fragments are electrophoretically pumped by means of an electric field of 170 V/cm applied across the two ends of the channel for a duration of 30–60s. The

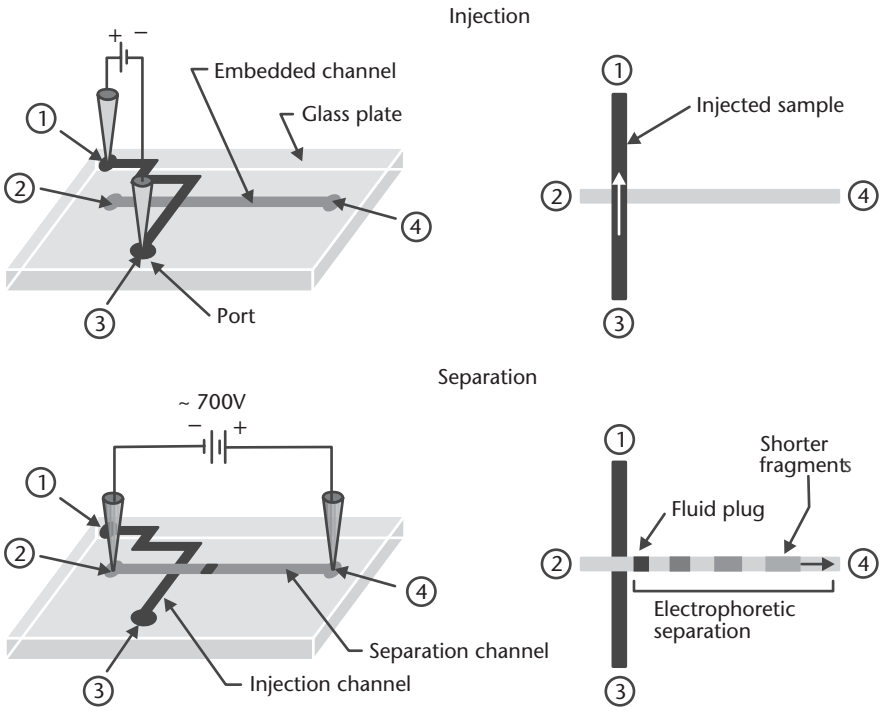


Figure 6.6 Illustration of the fluid injection and separation steps in a miniature DNA electrophoresis system. An applied electric field electrophoretically pumps the fluid molecules from port 3 to port 1 during the injection step. Another applied voltage between ports 2 and 4 initiates the electrophoretic separation of the DNA molecules. The smearing of the fluid plug in the separation channel is schematically illustrated. The capillary channels have a typical cross section of 8 × 50 μm^2. The separation capillary is 3.5 cm long. (*After:* [17, 18].)

injection-channel loading time is critical: If it is too short, more short DNA fragments are injected in the next step; if it is too long, the sample is biased toward longer fragments. Once the injection channel is filled, the applied voltage is switched to be across the two ends of the separation channel. The applied electric field directs the small "plug" of ionized fragments from the intersection of the two channels into the separation channel. After a short injection time, the ends of the injection channel are made positive to pull ionized fragments still in the injection channel back from the junction with the separation channel; otherwise, injection would occur continuously. At an applied electric field of 180 V/cm, it takes approximately 2 min to complete the separation of the DNA fragments in the injected plug. This compares with 8 to 10 hours to complete an equivalent separation using conventional gel electrophoresis or 1 to 2 hours with conventional capillary electrophoresis. Optical imaging of a fluorescent tag on each DNA fragment is used to detect the separated products inside the channel. The results from Woolley and Mathies indicate a resolution of a single nucleotide in DNA strands that are up to 1,000 nucleotides long.

Though this demonstration is an important accomplishment, much remains to be done before portable DNA sequencing instruments are available on the market. A complete sequencing system must integrate PCR with electrophoresis—or some

other DNA detection method—as well as include all fluid preparation and handling functions, such as pumping, valving, filtering, mixing of reagents, and rinsing. This demands the development of a complete system with many enabling technologies, MEMS being only one of them.

The electrophoresis part of the DNA sequencing process has been commercialized by companies such as Caliper Life Sciences, Inc., of Hopkinton, Massachusetts, with the product now being sold as the LabChip by Agilent Technologies. Up to 12 samples containing variable-length sections of DNA are placed in the disposable LabChip, which is inserted into the Agilent 2100 Bioanalyzer system for analysis. This system is about the size of a small suitcase, has a separate computer for control and data acquisition, and is powered by a wall outlet, making the system semiportable.

The entire LabChip structure is made of sheets of glass. Patterning of glasses is limited to usually photolithography and etching or laser ablation (see Chapter 3). The layers are bonded together under heat and pressure, then cut apart. The use of glass in a simple process leads to low cost, making a single use before disposal economical. Single-use devices have the advantages of no concern about cross contamination from previous samples, greatly reduced chances of clogging, and no long-term risk of material degradation with use. Many glasses (and plastics) are transparent to visible and UV light, which is useful in optical detection schemes. Some specifications for the Agilent DNA 1000 LabChip include a DNA concentration range of 0.5–50 ng/μl, a sizing range of 25–1,000 base pairs, a sizing accuracy of $\pm$15%, and a resolution better than 10% over most of the range [19]. The sample volume is 1 μl and takes 30 min to analyze.

DNA Hybridization Arrays

Once fragments of an unknown DNA sample have been amplified into many copies, they can be read with DNA hybridization arrays. These are different sequences of preassembled nucleotides attached to a substrate (see Figure 6.7). The DNA sections to be identified, with lengths in the range of a hundred to thousands of bases, are tagged with a fluorescent dye at one end. When placed in a buffer solution on the substrate, sections of some of the unknowns hybridize to the complementary sequences on the substrate. As discussed earlier, hybridization is the process by which DNA strands match up and bind with complementary DNA capture probes. The substrate is then rinsed and illuminated. The locations of fluorescence indicate where hybridization occurred and thus which sequences are present in the unknown. This approach is particularly beneficial in the detection of specific gene mutations and in the search for known pathogens.

Several companies commercially produce microscale DNA arrays. One of the market-leading products is the GeneChip® from Affymetrix of Santa Clara, California [20]. The GeneChip is produced on 5-inch square fused quartz substrates, which are coated with a bonding layer comprised of molecules to which the DNA nucleotides can adhere, followed by a *protection group* [21, 22]. Using a standard photolithographic mask (see Chapter 3), ultraviolet light is shone through 20-μm square openings to remove the protection groups, activating selected sites on the substrate (see Figure 6.8). A solution containing one type of nucleotide (A, T, G, or C) with a removable protection group is flushed across the surface. These

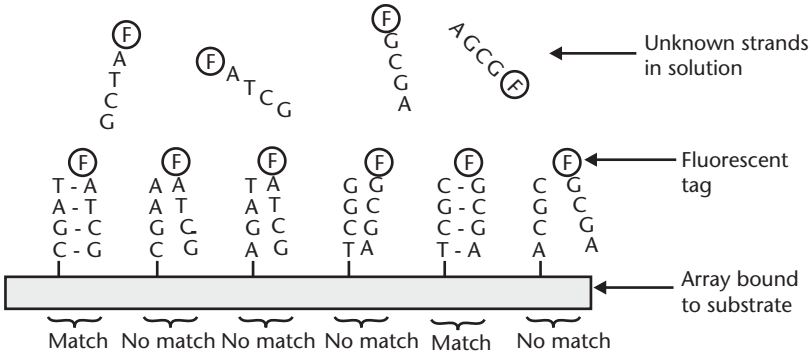


Figure 6.7 The use of a DNA hybridization array. Only complementary DNA fragments in the solution match can hybridize to the fragments bound to the substrate. The free fragments, which are usually much longer than the bound fragments, have fluorescent tags on the end for reading. Only sites that receive their complements will fluoresce when read.

nucleotides bond to activated sites in each square that was exposed but not in the other areas. The process is repeated to start chains of the other three-nucleotide types. Repeated exposure with different masks to remove the protection groups and flushing with the four nucleotide solutions grow DNA strands, or probes, that are

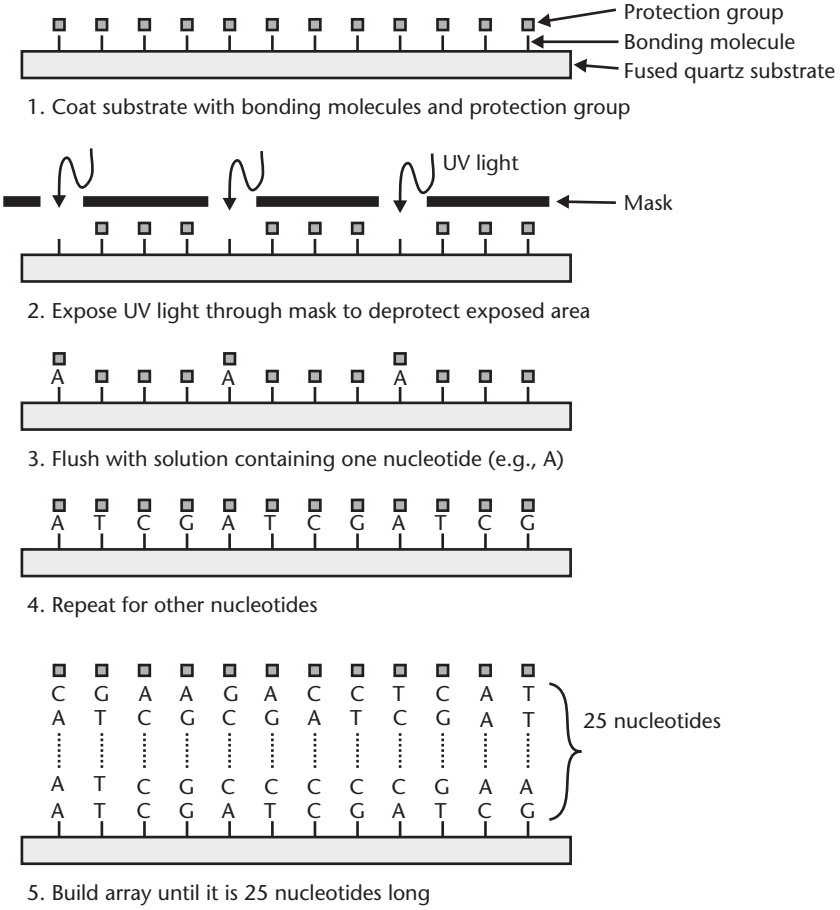


Figure 6.8 Illustration of the GeneChip fabrication process. (*After:* [20].)

typically 25 nucleotides long. Finally, all probes are deprotected, the substrates are diced, and they are packaged in plastic flow-cell cartridges for use.

With 25 nucleotides in a sequence, there are 4^{25} (equal to 10^{15}) different combinations that can be made with this process. However, with a final chip size of 1.28 cm^2, there is only enough space for about 320,000 squares with different sequences. Thus Affymetrix produces chips with only preselected sequences, targeting specific applications (e.g., detecting strains of *E. coli* or hereditary neurological disorders in humans). If different sequences or longer lengths are desired, custom arrays can be made either with a new mask set or with a special maskless project system, such as one based on Texas Instruments' DLP (see Chapter 5), available from BioAutomation of Plano, Texas [21].

Another microarray market leader is Agilent Technologies. One product, the Human 1A Oligo Microarray, has over 18,000 probes per 1- by 3-in glass slide with lengths of 60 nucleotides [23]. Agilent uses inkjet technology (see Chapter 4) to write the probes, base by base, with processing similar to that for the Affymetrix probes. Picoliter volumes of nucleotide "ink" write round spots approximately 130 μm across. In addition to standard products, custom arrays can be produced with a shorter turnaround time than with the masking production method. Agilent also manufactures the Microarray Scanner for reading the arrays and producing computer output. The large quantity of data produced by DNA analyses has spawned a new field of study termed *bioinformatics*, which seeks to develop algorithms to handle large genetic databases.

Microelectrode Arrays

Electrodes are extremely useful in the sensing of biological and electrochemical potentials. In medicine, electrodes are commonly used to measure bioelectric signals generated by muscle or nerve cells. In electrochemistry, electric current from one or many electrodes can significantly alter the properties of a chemical reaction. It is natural that miniaturization of electrodes is sought in these fields, especially for applications where size is important or arrays of electrodes can enable new scientific knowledge. Academic research on microelectrodes abounds. The reader will find a comprehensive review of microelectrodes and their properties in a book chapter by Kovacs [24].

In simple terms, the metal microelectrode is merely an intermediate element that facilitates the transfer of electrons between an electrical circuit and an ionic solution. Two competing chemical processes, *oxidation* and *reduction*, determine the equilibrium conditions at the interface between the metal and the ionic solution. Under oxidation, the electrode loses electrons to the solution; reduction is the exact opposite process. In steady state, an equilibrium between these two reactions gives rise to an interfacial space charge region—an area depleted of any mobile charges (electrons or ions)—separating a surface sheet of electrons in the metal electrode from a layer of positive ions in the solution. This is similar to the depletion layer at the junction of a semiconductor *p-n* diode. The interfacial space charge region is extremely thin, on the order of 0.5 nm, resulting in a large capacitance on the order of 10^{-5} F per cm^2 of electrode area. Incidentally, this is precisely the principle of

operation in electrolytic capacitors. A simple electrical model for the microelectrode consists of a capacitor in series with a small resistor that reflects the resistance of the electrolyte in the vicinity.

The fabrication of microelectrode arrays first involves the deposition of an insulating layer, typically silicon dioxide, on a silicon substrate (see Figure 6.9). Alternatively, an insulating glass substrate is equally suitable. A thin metal film is sputtered or evaporated and then patterned to define the electrical interconnects and electrodes. Gold, iridium, and platinum, being very chemically inert, are excellent choices for measuring biopotentials as well as for electrochemistry. Silver is also important in electrochemistry because many published electrochemical potentials are referred to silver/silver-chloride electrode. It should be noted that wire bonding to platinum or iridium is very difficult. If the microelectrode must be made of such metals, it is necessary to deposit an additional layer of gold over the bond pads for wire bonding. The deposition of a silicon nitride layer seals and protects the metal structures. Openings in this layer define the microelectrodes and the bond pads. The following sections describe two instances where microelectrodes show promise as a diagnostics tool in biochemistry and biology.

DNA Addressing with Microelectrodes

A unique and novel application patented by Nanogen of San Diego, California [25], makes use of microelectrode arrays in the analysis of DNA fragments of unknown sequences. The approach exploits the polar property of DNA molecules to attract them to positively charged microelectrodes in an array. The analysis consists of two sequential operations, beginning first with building an array of known DNA capture probes over the electrode array, followed by hybridization of the unknown DNA fragments. DNA capture probes are synthetic short chains of nucleotides of known specific sequence.

Applying a positive voltage to a selection of microelectrodes in the array attracts previously synthesized DNA capture probes to these biased electrodes, where they chemically bind in permeable hydrogel layer that had been impregnated with a coupling agent (see Figure 6.10) [26]. Microelectrodes in the array that are negatively biased remain clear. Subsequent washing removes only unbound probes. Immersion

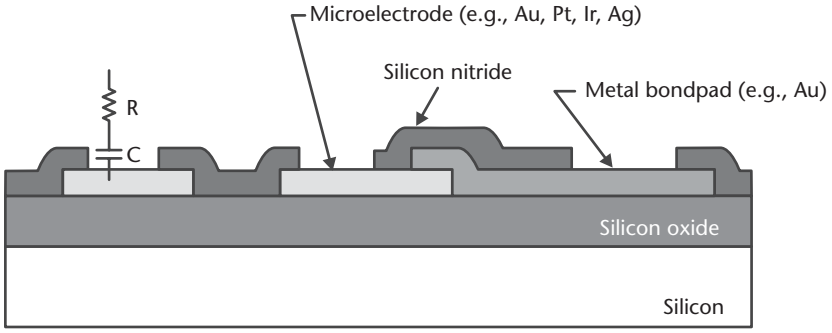


Figure 6.9 Cross section of a microelectrode array showing two different metals for the electrodes and for the bond pads. The schematic also illustrates a basic electrical equivalent circuit that emphasizes the capacitive behavior of a microelectrode. The silicon substrate and the silicon dioxide dielectric layer may be substituted by an insulating glass substrate.

in a second solution binds a second type of DNA capture probes to another set of biased electrodes. Repetition of the cycle with appropriate electrode biasing sequentially builds a large array containing tens and potentially hundreds of individually distinct sites of DNA capture probes differing by their sequence of nucleotides. The removal of a capture probe from a particular site, if necessary, is simple, accomplished by applying a negative potential to the desired microelectrode and releasing the probe back into the solution. It is this electrical addressing scheme to selectively attract or repel DNA molecules that makes this method versatile and powerful.

Once the array of DNA capture probes is ready, a sample solution containing DNA fragments of unknown sequence (target DNA) is introduced. These fragments hybridize with the DNA capture probes—in other words, the target DNA binds only to DNA capture probes containing a complementary sequence. Optical imaging of fluorescent tags reveals the hybridized probe sites in the array and, consequently, information on the sequence of nucleotides in the target DNA. This approach is particularly beneficial in the detection of specific gene mutations or in the search for known pathogens.

Positive biasing of select electrodes during the hybridization phase accelerates the process by actively steering and concentrating with the applied electric field target DNA molecules onto desired electrodes. Accelerated hybridization occurs in minutes rather than the hours typical of passive hybridization techniques. The

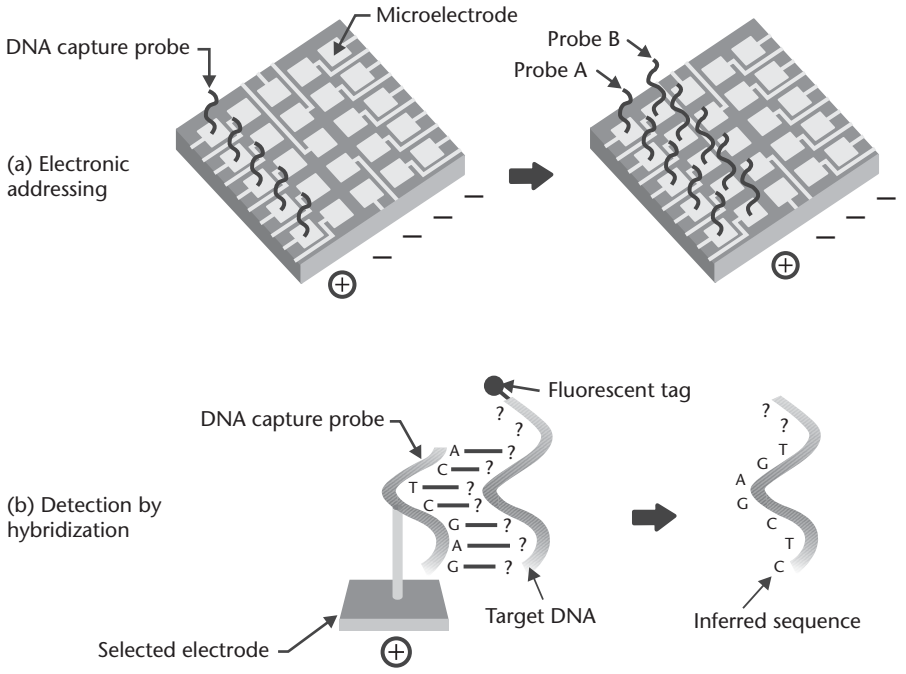


Figure 6.10 Illustration of the Nanogen electronic addressing and detection schemes. (a) A positive voltage attracts DNA capture probes to biased microelectrodes. Negatively biased electrodes remain clear of DNA. Repetition of the cycle in different solutions with appropriate electrode biasing sequentially builds an array of individually distinct sites of DNA capture probes that differ by their sequence of nucleotides. (b) A DNA fragment with unknown sequence hybridizes with a DNA capture probe with a complementary sequence. Fluorescence microscopy reveals the hybridized site and, consequently, the unknown sequence.

method is sufficiently sensitive to detect single base differences and single-point mutations in the DNA sequence.

Cell Cultures over Microelectrodes

Many types of cells, in particular nerve and heart cells, can grow in an artificial culture over a microelectrode array. The growth normally requires a constant temperature, often at 37°C (the core temperature of the human body), a suitable flow of oxygen, and a continuous supply of nutrients [27]. Bioelectric activity, or action potential, capacitively couples across the cell membrane and surrounding fluid to the nearest microelectrode, which then measures a small ac potential, typically between 10 and 1,000 μV in peak amplitude. The array of microelectrodes essentially images the dynamic electrical activity across a large sheet of living cells. The measured action potentials and their corresponding temporal waveforms are characteristic of the cell type and the overall health of the cell culture. For example, toxins that block the flow of sodium or potassium ions across the cell membrane suppress the action potentials or alter their frequency content (see Figure 6.11) [27]. This approach may be useful in the future for studying the effects of experimental drugs *in vitro* or for the early detection of airborne toxic particles.

Summary

In recent years, a number of microscale biological analysis techniques have become commercialized, notably electrophoresis and arrays for DNA analysis on disposable glass or plastic chips. Prototypes and products to run analyses are becoming smaller and more portable. Most of these biological applications employ microfluidics, in which pumping methods are different than in the macroscopic world and Reynolds numbers are very low.

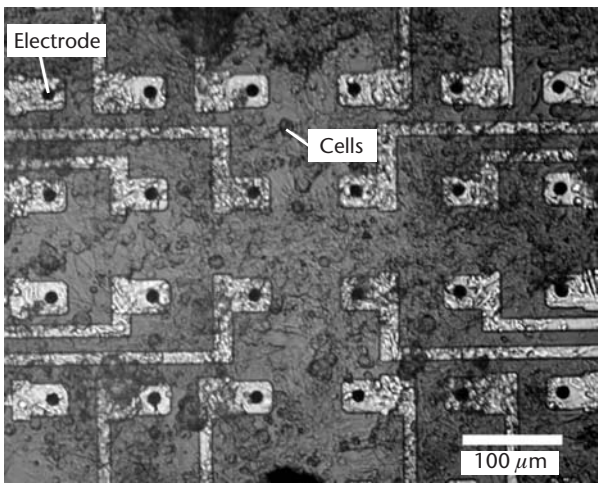


Figure 6.11 Photograph of a cultured *syncytium* spontaneously beating over a microelectrode array. The platinum electrodes are 10 μm in diameter with a spacing of 100 μm. The electrodes measure the extracellular currents generated by a traveling wave of action potential across the sheet of living cells. (*Courtesy of:* B. D. DeBusschere of Stanford University, Stanford, California.)

References

[1] Manz, A., N. Graber, and H. M. Widmer, "Miniaturized Total Chemical Analysis Systems: A Novel Concept for Chemical Sensing," *Sensors and Actuators B*, Vol. B1, 1990, pp. 244–248.

[2] Kovacs, G. T. A., *Micromachined Transducer Sourcebook*, Boston, MA: WCB McGraw-Hill, 1998, Section 6.6.

[3] Sharp, K. V., et al., "Liquid Flow in Microchannels," in *The MEMS Handbook*, M. Gad-el-Hak (ed.), Boca Raton, FL: CRC Press, 2002, Chapter 6.

[4] Kopf-Sill, A. R., et al., "Creating a Lab-on-a-Chip with Microfluidic Technologies," in *Integrated Microfabricated Biodevices*, M. J. Heller and A. Guttman (eds.), New York: Marcel Dekker, 2002, Chapter 2.

[5] Gray, B. L., et al., "Novel Interconnection Technologies for Integrated Microfluidic Systems," *Sensors and Actuators A*, Vol. 77, 1999, pp. 57–65.

[6] Stryer, L., *Biochemistry*, New York: W. H. Freeman and Co., 1988, pp. 71–90, 120–123.

[7] Darnell, J., L. Harvey, and D. Baltimore, *Molecular Cell Biology*, 2nd ed., New York: Scientific American Books, 1990, p. 219.

[8] Nguyen, N. -T., and S. T. Wereley, *Fundamentals and Applications of Microfluidics*, Norwood, MA: Artech House, 2002.

[9] Mastrangelo, C. H., M. A. Burns, and D. T. Burke, "Microfabricated Devices for Genetic Diagnostics," *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1769–1787.

[10] Wilding, P., M. A. Shoffner, and L. J. Kricka, *Clinical Chemistry*, Vol. 40, No. 9, September 1994, pp.1815–1818.

[11] U. S. Patent 5,674,742, October 7, 1997.

[12] Northrup, M. A., et al., "DNA Amplification with a Microfabricated Reaction Chamber," *Proc. 7th Int. Conf. on Solid-State Sensors and Actuators*, Yokohama, Japan, June 7–10, 1993, pp. 924–926.

[13] Belgrader, P., et al., "Development of Battery-Powered, Portable Instrumentation for Rapid PCR Analysis," in *Integrated Microfabricated Biodevices*, M. J. Heller and A. Guttman (eds.), New York: Marcel Dekker, 2002, Chapter 8.

[14] TaqMan® EZ-RT PCR Kit, Protocol, Applied Biosystems, Foster City, CA, 2002.

[15] Kuhr, W. G., and C. A. Monnig, "Capillary Electrophoresis," *Analytical Chemistry*, Vol. 64, 1992, pp. 389R–407R.

[16] Manz, A., et al., "Planar Chips Technology for Miniaturization and Integration of Separation Techniques into Monitoring Systems. Capillary Electrophoresis on a Chip," *Journal of Chromatography*, Vol. 593, 1992, pp. 253–258.

[17] Woolley, A. T., and R. A. Mathies, "Ultra-High Speed DNA Sequencing Using Capillary Electrophoresis Chips," *Analytical Chemistry*, Vol. 67, 1995, pp. 3676–3680.

[18] Woolley, A. T., and R. A. Mathies, "Ultra-High Speed DNA Fragment Separations Using Capillary Array Electrophoresis Chips," *Proceedings of the National Academy of Sciences USA*, Vol. 91, November 1994, pp. 11348–11352.

[19] Agilent Technologies, Product Literature for DNA1000 LabChip Kit, Palo Alto, CA, 2001.

[20] Affymetrix, GeneChip Product Literature, Santa Clara, CA, 2003.

[21] Garner, H. R., R. P. Balog, and K. J. Luebke, "Engineering in Genomics," *IEEE Engineering in Medicine and Biology*, July/August 2002, pp. 123–125.

[22] Fodor, S. P., et al., "Multiplexed Biochemical Assays with Biological Chips," *Nature*, Vol. 364, No. 6437, 1993, pp. 555–556.

[23] Agilent Technologies, Inc., Product Brochure for Agilent SurePrint Technology, Palo Alto, CA, 2001.

[24] Kovacs, G. T. A., "Introduction to the Theory, Design, and Modeling of Thin-Film Microelectrodes for Neural Interfaces," in *Enabling Technologies for Cultured Neural Networks*, D. A. Stenger and T. M. McKenna (eds.), San Diego, CA: Academic Press, 1994, pp. 121–166.

[25] U.S. Patents 5,605,662, February 25, 1997, and 5,632,957, May 27, 1997.

[26] Heller, M. J., et al., "Active Microelectronic Array Systems for DNA Hybridization, Genotyping, Pharmacogenomic, and Nanofabrication Applications," in *Integrated Microfabricated Biodevices*, M. J. Heller and A. Guttman (eds.), New York: Marcel Dekker, 2002, Chapter 10.

[27] Borkholder, D. A., B. D. DeBusschere, and G. T. A. Kovacs, "An Approach to the Classification of Unknown Biological Agents with Cell Based Sensors," *Tech. Digest Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, June 8–11, 1998, pp. 178–182.

Selected Bibliography

Heller, M. J., and A. Guttman (eds.), *Integrated Microfabricated Biodevices*, New York: Marcel Dekker, 2002.

Horton, R. M., and R. C. Tait, *Genetic Engineering with PCR*, Norfolk, UK: Horizon Press, 1998.

The reader will find extensive coverage of the research activities in this field in past proceedings of the conference on Micro Total Analysis Systems (μTAS).

CHAPTER 7

MEM Structures and Systems in RF Applications

> "The discovery of electrical waves has not merely scientific interest though that alone inspired it... it has had a profound influence on civilization; it has been instrumental in providing the methods which may bring all inhabitants of the world within hearing distance of each other and has potentialities social, educational and political which we are only beginning to realize."
>
> —*Sir Joseph. J. Thomson, on James Maxwell's discovery of electromagnetic waves in* James Clerk Maxwell: A Commemorative Volume 1831–1931*, The University Press: Cambridge, UK, 1931.*

Radio-frequency (RF) MEM devices have been in research and development for years, with scores of papers published annually. There are unpublicized devices in use in small volume in commercial and military applications, but only recently have such devices gone into high-volume production. Current and future RF MEMS devices will be competitive with more conventional components on the basis of volume, mass, cost, and performance. The largest potential market is in cellular telephone handsets, with hundreds of millions of units sold each year. Other portable electronics markets, where the aforementioned qualities are major considerations, include cordless phones for home use, wireless computer networking, radios, and global positioning system (GPS) receivers. Satellites, missile guidance, military radar, and test equipment are separate markets of importance, with lower potential sales volumes but higher unit prices.

Opening the cover of a modern cellular telephone reveals a myriad of discrete passive and active components occupying substantial volume and weight. The market's continued push for small portable telephones argues a convincing economic case for the miniaturization of components. MEMS technology promises to deliver miniature integrated solutions including variable capacitors, inductors, oscillators, filters, and switches to potentially replace conventional discrete components.

Signal Integrity in RF MEMS

A requirement for any RF device is maintaining signal integrity: transmitting desired signals with low loss, minimizing reflections, not permitting external signals or noise to join the transmitted signal, and filtering out or not generating undesired signals, such as higher-frequency harmonics. At high frequencies, these seemingly simple requirements are not readily attained.

Basic electromagnetic theory teaches that when the signal wavelength is on the order of the size of the system through which it flows, it is necessary to use a specialized type of electrical connections called transmission lines to carry the electrical signal from one point to another [1]. Transmission lines have a conductor for the signal and one or multiple nearby ground conducting lines running parallel to the signal line. The familiar coaxial lines used for cable television are one example where the signal conductor is in the center of a hollow cylindrical ground conductor. On circuit boards, strip lines, which have a signal line sandwiched between two ground planes and separated by a dielectric, are common. Another form that is more easily implemented on circuit boards, and especially on chips, is the coplanar waveguide. This has a central strip of metal for the signal, with ground strips on both sides. The whole structure resides on a dielectric, with air or vacuum above. Devices that transmit the RF signal, such as switches, must match the characteristic impedance of the transmission line to avoid signal reflections. Similarly, devices at the output terminals of a transmission line must be impedance-matched to collect the full signal strength and avoid undesirable reflection.

Losses fall into two categories: conductor loss and dielectric loss. Conductor loss is due to the nonzero resistance of the materials used, resulting in heating, and is modeled as an equivalent resistance *in series* with the signal path. Low-resistivity metals such as gold, copper, and aluminum are therefore commonly used for the conductors in RF MEMS. Contacts in a switch and even between layers of different materials also add resistance, which must be considered. Dielectric loss is due to atomic-scale dipoles excited in the dielectric material, also resulting in heating, and can be modeled as an equivalent *parallel* conductance. Eddy currents induced in an underlying conducting substrate are also modeled as a parallel conductance. Insulating or semiinsulating substrates are often necessary for RF-MEMS devices to minimize the loss due to eddy currents. High-resistivity (>5,000 Ω•cm) silicon substrates are acceptable but semi-insulating gallium arsenide (GaAs) is preferred when possible. Gallium arsenide is already in common usage for microwave integrated circuits for its high electron mobility. Insulators such as glass and alumina are preferred from a low-loss standpoint, although other considerations such as process compatibility, cost, and thermal coefficient of expansion mismatch factor into substrate selection.

Passive Electrical Components: Capacitors and Inductors

Quality Factor and Parasitics in Passive Components

All capacitors and inductors have parasitics associated with them that limit their performance. Two parameters that describe their performance and enable comparisons between devices are the quality factor Q and the self-resonance frequency f_{SR}.

The quality factor Q is a measure of loss in a linear-circuit element and is defined as the maximum energy stored during a cycle divided by the energy lost per cycle. For reactive components such as capacitors and inductors, it is equal to the absolute value of the ratio of the imaginary part of the impedance to the real part of the impedance: for a capacitor C with series resistance R_s, $Q = 1/(2\pi fCR_S)$; for an inductor L with series resistance R_s, $Q = 2\pi fL/R_S$ [see Figure 7.1 (a, b)]. In both cases, a

greater resistance gives a smaller Q. As the frequency goes up, current flows along an increasingly thin layer at the surface of conductors (the skin depth) [1], increasing the resistance and lowering the Q. For capacitors, dielectric loss, which is a function of frequency [2], also contributes to a lower Q. For micromachined capacitors, however, the dielectric is usually a gas or vacuum, leaving series resistance as the dominant loss mechanism in many designs; however, loss can also occur in the substrate [see Figure 7.1(a)]. Because of the long, thin shape of their conductors, inductors tend to have a higher series resistance than do capacitors, resulting in far lower quality factors. On-chip inductors can have substrate loss as well [see Figure 7.1(b)]. It is clear from the equations that quality factor for a given component varies with frequency and the value of capacitance or inductance, so both frequency and component value must be specified when citing a value for Q and especially when making comparisons in order to be fair.

The lines connecting a capacitor to a circuit and even the plates themselves have a small parasitic inductance [1]. A circuit model has a capacitance C in series with an inductance L_{para} [see Figure 7.1(c)]. At low frequencies, the impedance of the inductor, which is imaginary and positive, is small, and the capacitor functions as normal. As the frequency rises, however, the capacitor impedance, which is imaginary and negative, falls while the impedance of the inductor rises, limiting the useful

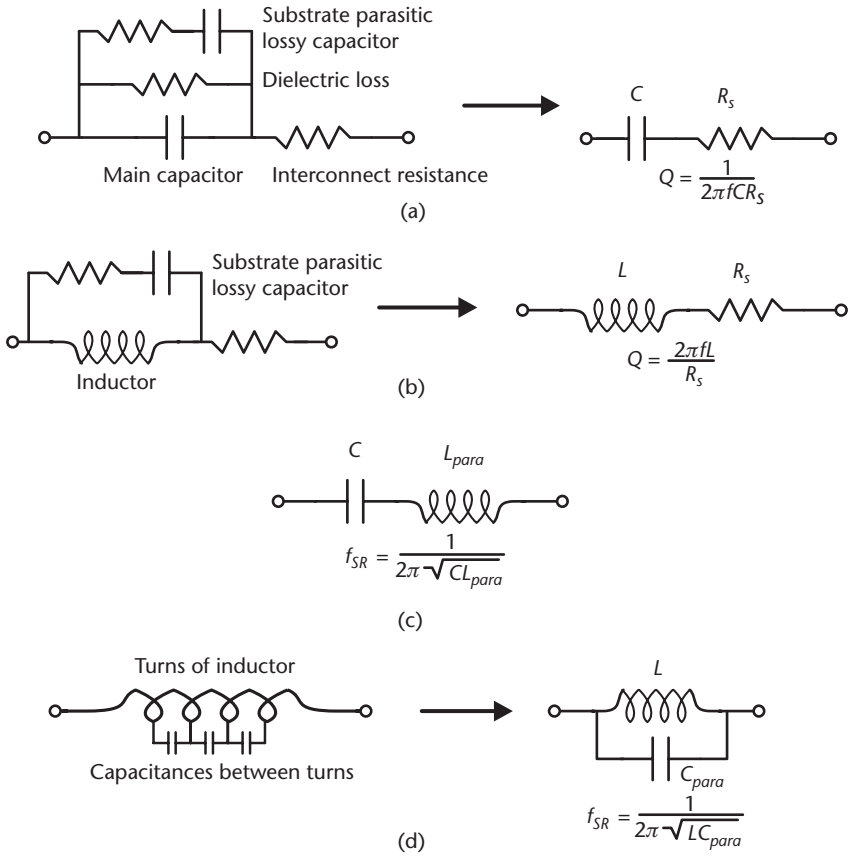


Figure 7.1 Parasitics in reactive devices: (a) micromachined capacitor modeled as a capacitor with parasitic series resistance; (b) inductor with parasitic series resistance; (c) capacitor with parasitic series inductance; and (d) inductor with parasitic parallel capacitance between coils.

frequency range. Above the self-resonance frequency $f_{SR} = 1/\left(2\pi\sqrt{CL_{para}}\right)$, the inductance dominates and the capacitor looks to a circuit like an inductor (i.e., the pair has an imaginary positive impedance).

Inductors are usually implemented as coils of a conductor, which have parasitic capacitance between them. A circuit model can be made with each turn of the inductor represented by an incremental inductor and its parasitic capacitance [see Figure 7.1(d)]. A simplified model has an inductor L in parallel with a parasitic capacitor C_{para}. At low frequencies, the capacitor has a large imaginary negative impedance, and most current flows through the inductor. As the frequency rises, however, the magnitude of the capacitor impedance falls, while the imaginary positive impedance of the inductor rises. Above the self-resonant frequency $f_{SR} = 1/\left(2\pi\sqrt{LC_{para}}\right)$, the capacitance dominates (i.e., the pair has an imaginary negative impedance), and the inductor ceases to function as one. In general, this occurs at a lower frequency for inductors than for capacitors. As seen in the equation for Q, the quality factor for an inductor rises with frequency; however, because the parallel capacitance reduces the effective inductance at higher frequencies, Q eventually reaches a maximum before falling.

Surface-Micromachined Variable Capacitors

Capacitors with a constant capacitance are readily fabricated side by side with transistors in standard semiconductor integrated-circuit processes by sandwiching a dielectric between two conductive layers. The primary reason for using on-chip capacitors is the reduction in parts that must be used on a circuit board and the commensurate reduction in cost. Other reasons include noise reduction and lowering both parasitic capacitance and resistance. Because the capacitance per unit area in a standard process is relatively small, large capacitances (more than a few picofarads) occupy too great a chip area to be cost effective, and high-dielectric materials must be integrated or off-chip components must be used.

Some analog circuits, such as voltage-controlled oscillators (VCOs) and tuning circuits, require voltage-controlled variable capacitors (varactors). These are presently implemented on a separate semiconductor chip with a reverse-biased *p-n* diode junction. Varying the dc voltage applied varies the depletion-region width and thus the small-signal capacitance; capacitance tuning ranges from 2:1 to over 10:1 with an applied voltage of 0–5V are available commercially. The greatest limitation of semiconductor varactors is their Q, which is, at most, on the order of 50 in the gigahertz frequency range. High Q is required for oscillators with low phase noise [3]; an example requirement is $Q > 50$ for a 2-pF capacitor in the range of 1 GHz [4]. Micromachining technology is expected to make an impact in the near future with the commercial fabrication of variable capacitors with higher Q, the ability to be fabricated on the same chip as semiconductor circuitry for a reduction in part count, the ability to handle large ac input voltages that would forward bias diode varactors, and potentially wider tuning range.

Micromachined variable capacitors can be divided into two broad categories, surface-micromachined and bulk-micromachined. Surface-micromachined variable capacitors tend to be simpler to fabricate, more readily integrated on the same chip as existing circuitry, and use less expensive process steps than their

bulk-micromachined counterparts, but they have a nonlinear response to the tuning voltage and smaller tuning ranges. The quality factor and self-resonance frequency vary with the design.

Many versions of surface-micromachined variable capacitors have been demonstrated in research papers and patents [5–7]. Most implementations have in common a bottom plate residing on an insulated substrate, an air gap, and a flat top plate parallel to the substrate suspended by a spring structure [see Figure 7.2(a)]. Applying a dc control voltage V creates an electrostatic force $F_e = \varepsilon_0 A V^2/(2g^2)$, where ε_0 is the permittivity of free space, A is the area of plate overlap, and g is the gap. This force pulls the top plate downward, increasing the capacitance. The restoring spring force is given by $F_s = k\Delta g$, where k is the spring constant and Δg is the plate motion or displacement. The spring force increases linearly with plate motion, but the electrostatic force rises faster than linearly with the plate gap change. This results in both the plate motion and the capacitance changing slowly at first, then rising rapidly. When the displacement reaches one third of the initial gap, the electrostatic force rises more rapidly than the spring force, and the top plate snaps down toward the bottom plate. This limits the controllable increase in capacitance for this type of variable capacitor to 50%, which is sufficient for many VCO applications. Parasitic capacitance, which does not change with voltage, lowers the possible tuning range.

In portable applications such as cellular-phone handsets, the dc control voltage is limited to 3.6V or less (on-chip charge-pump circuitry can, however, increase the available voltage). A system-determined capacitance and process-determined gap set the required mechanical spring constant. Another design consideration is that electrical current must flow through the springs to the top plate, making the springs the dominant source of series resistance. The geometrical dimensions of the springs can be optimized to provide the least electrical resistance for a particular spring constant.

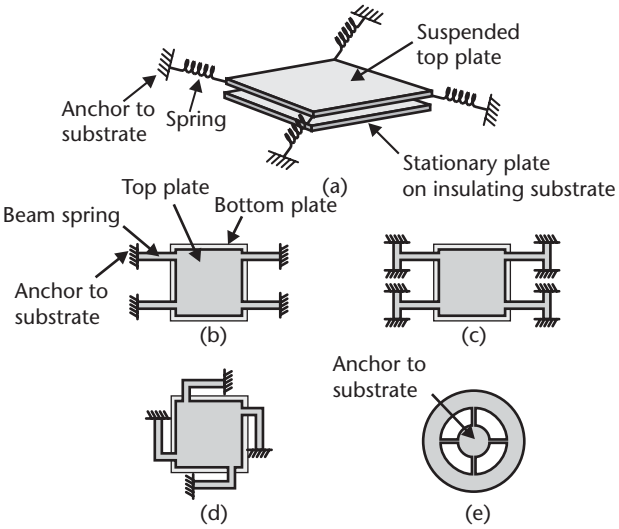


Figure 7.2 Different implementations of a surface-micromachined parallel-plate variable capacitor: (a) perspective view of the basic concept showing a stationary plate and a moveable plate suspended by springs; (b) top view of a capacitor using straight beams as springs; (c) top view using T-shaped springs [8]; (d) top view using L-shaped springs [5]; and (e) top view with center anchor [7].

An optimal solution is to make the spring beams short, thick, and wide, but within the constraint of the spring constant. The resistance can also be kept to a minimum by the use of a highly conductive metal.

Design and fabrication so that the top plate moves as desired present a particular challenge due to the effects of stress. Similar challenges are faced by other freestanding micromachined structures, making this a subject worth exploring. The use of straight beam springs, such as those shown in Figure 7.2(b), does not allow stress relief. Compressive residual stress can cause the plate to flex upward or downward, while tensile residual stress increases the voltage that must be applied to cause a given amount of plate motion. Even if the plate material is deposited with a low residual stress, differential thermal expansion between the freestanding structure and the substrate results in stress that varies with temperature (e.g., many consumer products are specified to operate over wide range from –20° to +40°C). Differential thermal expansion problems for this type of design can be avoided by choosing substrate and structural materials with the same coefficient of thermal expansion, such as single-crystal silicon and polycrystalline silicon. Because silicon is neither a good insulator nor a good conductor, this approach is not adequate if the design requires an insulating substrate and a highly conducting capacitor to minimize losses, as discussed earlier.

Several different designs have been made in attempts to reduce the effects of stress and simultaneously achieve a high Q. The design shown in Figure 7.2(c) uses LPCVD polysilicon springs and top plate, with the beam springs laid out in a T shape in an attempt to relieve stress, but fabricated structures were still warped at room temperature [8]. A thin layer of gold on top of the polysilicon lowers the resistance, giving a Q of 20 at 1 GHz and a self-resonant frequency above 6 GHz for a 2-pF capacitor. The measured tuning range is 1.5:1 at 4V. In general, a single material should be used if it is desired that a freestanding structure remain flat; the metal on silicon used in this capacitor imparts an undesired curvature to the plate as the temperature varies.

The design in Figure 7.2(d) uses sputtered aluminum for the top plate and springs, with a sacrificial photoresist layer that is removed in an oxygen plasma [5]. The circularly oriented L shape of these springs allows the plate to rotate to give some relief of both residual and temperature-induced stresses. Even with this design, however, plate warpage was sufficiently large that only capacitors smaller than 200 μm on a side could be used, so several small capacitors are wired in parallel. Capacitors with a nominal 2.1-pF value have a Q of 62 at 1 GHz and a tuning range of 1.16:1 at 5.5V. An aluminum ground plane under the bottom plate, traces, and bond pads shields the silicon substrate, reducing eddy-current loss. As there is also capacitance that does not vary with voltage between the traces and the ground, the maximum possible tuning range with a higher voltage is less than 1.5:1. Unlike the polysilicon-based capacitor described earlier, the use of low-temperature process steps allows integration of such capacitors on wafers with previously fabricated circuitry.

The design in Figure 7.2(e) [7] uses electroplated gold for the plate and springs, with a sacrificial PECVD oxide layer etched in hydrofluoric acid followed by supercritical drying (see Chapter 3). In contrast to the other designs, the anchor is in the center of the structure, which allows the outer edge of the plate to freely expand or contract in order to relieve both residual and thermal stresses. A bond wire to the

center makes an electrical contact with low resistance and low inductance. The greatest challenge in this design is fabricating the plate metal with a low residual stress *gradient* to avoid curling. For a nominal capacitance of 2 pF, this design has a Q of 181 at 1 GHz and a self-resonance frequency of 7.5 GHz. The tuning range is 1.45:1 at 5.5V. To demonstrate the deleterious effect of substrate conductivity, capacitors were fabricated with the same process on fused quartz and on high-resistivity (>10 kΩ•cm) silicon substrates with 4 μm of silicon dioxide for an insulator. The small parasitic conductance through the silicon reduced Q by a factor of 40 compared to that on quartz. The use of an insulating substrate also reduced the parasitic capacitance in the traces and bond pads, which was a concern with the other designs, to about 1% of the nominal capacitance.

All of the surface-micromachined designs described have small etch holes in the top plate to allow the etchant to remove the sacrificial layer rapidly. The last design also has a layer of silicon nitride coating the bottom plate to prevent shorting when the top plate snaps down. It further has standoff bumps protruding under the top plate to limit motion and reduce the contact area at snap down, reducing the likelihood of sticking.

Bulk-Micromachined Variable Capacitors

The most successful bulk-micromachined variable capacitors have been of the interdigitated-finger (comb-drive) type [see Figure 7.3(a)]. In these devices, a spring supports a set of movable fingers that mesh with a set of stationary fingers. When a dc voltage is applied, the electrostatic force attracts the movable fingers to increase

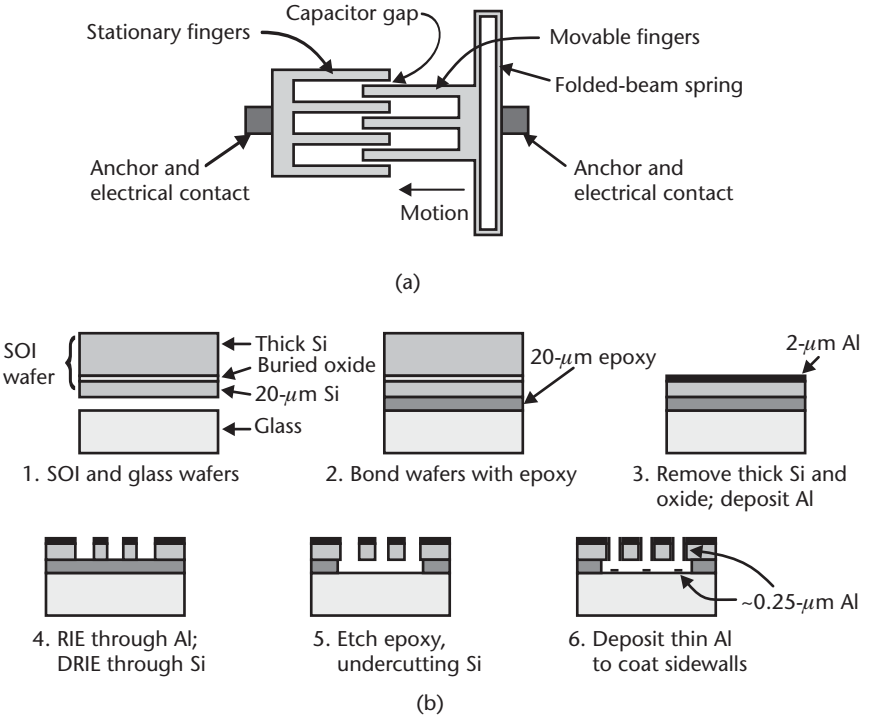


Figure 7.3 Interdigitated-finger capacitor: (a) conceptual top view showing a group of fingers and springs; and (b) process flow for reduced substrate parasitics. (*After:* [11].)

the length of overlap and thus the capacitance between the fingers. The capacitance scales linearly with the number of fingers and the finger thickness and is inversely proportional to the gap. Thus, the use of DRIE for high-aspect-ratio trenches results in a large capacitance per unit area. Fabricating a massive number of fingers in parallel, typically in several separate blocks, gives the desired capacitance. As with the surface-micromachined capacitors, resistance in the springs dominates the Q if other sources of loss are minimized.

Several generations of interdigitated-finger variable capacitors have been designed and fabricated at the Rockwell Science Center of Thousand Oaks, California [9–11]. The earliest was fabricated using a single mask to DRIE though the top layer of silicon on a SOI wafer, stopping on a 2-μm-thick buried layer of silicon dioxide. Hydrofluoric acid removes the buried oxide in the regions with moving parts, followed by supercritical drying. Even heavy doping ($>10^{20}$ cm^{-3}) of the silicon springs and fingers does not yield as low a sufficiently low resistance, so their resistance is lowered by sputtering on a thin, unpatterned layer of aluminum. This coats the tops and sidewalls of the silicon. The top of the underlying silicon substrate is also coated, but gaps are formed due to the undercut of the oxide, avoiding shorting. Even with the thick buried oxide, the underlying low-resistivity silicon handle substrate acts as the lossy plate of a capacitor, giving both low Q and a large parasitic capacitance [see Figure 7.1(a)]. Changing the underlying silicon substrate to high-resistivity material improves the Q and lowers the parasitic capacitance by a factor of ten.

Further improvement in Q is achieved by using a glass substrate, which requires major changes in the process flow [see Figure 7.3(b)]. These capacitors are fabricated by bonding the thin-silicon side of a SOI wafer to the glass wafer with 20 μm of epoxy. The Young's moduli of epoxies are normally orders of magnitude lower than those of silicon and glass, providing some stress isolation between the two substrates. The thick side of the SOI wafer is removed by grinding and polishing off most of the silicon, completing the silicon removal by etching in TMAH, and finally etching off the buried oxide to reveal the 20 μm-thick layer of silicon. Next, a 2-μm layer of aluminum is deposited for a low series resistance and patterned using standard lithography and plasma etching. The photoresist and aluminum act as a mask in a DRIE step to etch straight down through the silicon to the buried layer of epoxy. The epoxy is etched in an oxygen plasma, which undercuts the silicon to free the moving parts. An additional 0.25 μm of aluminum is sputtered to coat the sidewalls for an even lower series resistance. Fabricated capacitors have fingers that are 2 μm wide, 20 μm thick, a gap of 2 μm, and an initial overlap of a few micrometers. A nominal capacitance of 2 pF requires 1,200 sets of interdigitated fingers. The measured value for Q is 61 at 1 GHz, the self-resonance frequency is 5 GHz, and the tuning range is 4.55:1 at 5.2V from a finger motion of 23 μm.

A characteristic common to all mechanical devices with masses and springs, whether surface or bulk micromachined, or even macro- or microscale, is that the mass will move unintentionally when the device is subject to an external acceleration, such as vibration (this sensitivity results in noise that is casually referred to as *microphonics*). The displacement and corresponding capacitance change in variable capacitors depends on the acceleration, the mass, and the spring constant in the direction of motion. The allowed capacitance change is application-dependent.

A further concern for interdigitated-finger capacitors is motion of the movable fingers sideways, perpendicular to the intended direction of travel [along the vertical direction in Figure 7.3(a)]. In this event, the gap on one side is reduced, and the electrostatic force increases rapidly. This eventually pulls the fingers together, resulting in an electrical short. The spring constant perpendicular to the direction of travel must be sufficiently large to prevent any such displacement under the expected operating conditions. Microphonics is a key concern that must be resolved before micromachined variable capacitors are fully deployed in commercial systems.

Micromachined Inductors

Billions of low-cost discrete inductors are sold annually for applications including RF filters, VCOs, and chokes. Most of these have inductances in the range of a few to tens of nanohenries. When used in conjunction with an integrated circuit, discrete circuit components suffer from parasitic capacitance in traces and bond pads on the chip, in the bond wires connecting the chip to the circuit board, and in the inductor packaging. This limits the self-resonance frequency and therefore the maximum operating frequency. The inductors also consume precious board space in portable electronics; for example, in a Nokia 6161 cellular telephone, there are 24 discrete inductors (in addition to even more capacitors and resistors) along with only 15 integrated circuits [12]. To alleviate the self-resonance shortcoming and to reduce the part count and space used on a printed circuit board, low-cost, high-performance on-chip inductors are desirable.

Example inductor parameters needed for use in an on-chip high-Q resonant tank circuit for VCOs in cellular phones in the 1–2 GHz range are $L = 5$ nH and $Q > 30$ [4]. Inductors are readily fabricated on integrated-circuit chips using standard CMOS or bipolar processes by simply forming a spiral in one layer of metal and a connection to the center of the spiral in another layer of metal (see Figure 7.4). Losses from the resistance of the metal and eddy currents in the substrate limit the Q to less than 10 at 2 GHz [11].

One approach to improving both quality factor and self-resonance frequency is to reduce the parasitic capacitance and substrate conductive loss by changing to an insulating substrate, which is not possible if circuitry must be integrated on the same chip. Alternatively, raising the inductor above the substrate using an air gap or forming a cavity underneath it reduces the parasitic capacitance to the substrate. As an example, 24-nH inductors were made using a 12.5-turn spiral with an outer diameter of 137 μm. Those fabricated on the substrate have a self-resonance frequency of 1.8 GHz; those raised 250 μm above the substrate have an f_{SR} of 6.6 GHz [11]. The quality factor undoubtedly increased as well, but values were not reported. In a similar comparison, 1.2-nH inductors fabricated on the substrate with a f_{SR} of 22 GHz showed an increase to 70 GHz after substrate removal [3]. The quality factor for the latter was expected to be in the range of 60–80 at 40 GHz.

Another obvious solution for improvement in Q is minimizing the resistance by using a thick layer (limited by the skin depth) of low-resistivity metal. While integrated circuit–process inductors have been limited to the metal available in the process (usually aluminum), when given the choice of metals, researchers have chosen primarily copper and gold. A further improvement that may not be immediately

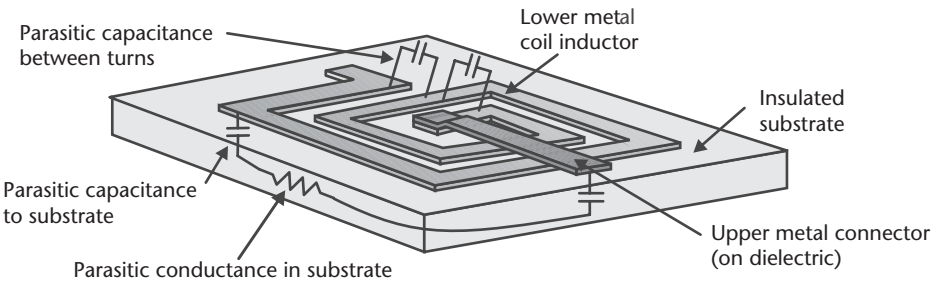
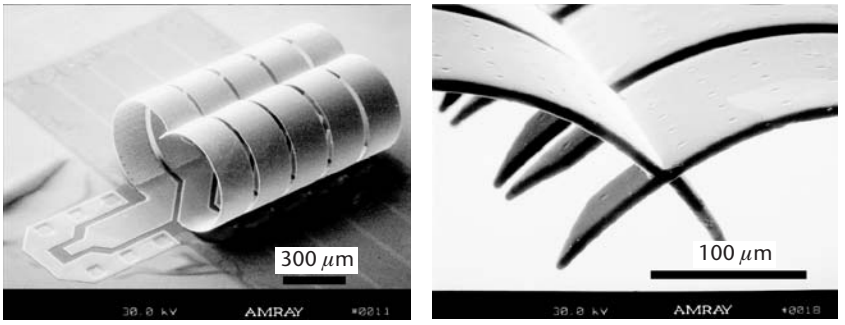


Figure 7.4 Illustration of a planar on-chip inductor, with parasitics noted. The inductor consists of a planar spiral made in one layer of metal and a connection to the center of the spiral in another layer of metal.

apparent is the addition of a highly conductive ground plane under the coil. Eddy currents are still induced, but the losses are much lower than with a resistive material.

An alternative to planar spiral inductors is positioning small solenoids on top of the substrate. The Palo Alto Research Center (PARC) of Palo Alto, California, is commercializing this concept [13]. In the PARC implementation, the coils are closely spaced ribbons of copper-plated metal that concentrate most of the magnetic flux inside the coils [see Figure 7.5(a)]. The bottoms of the coils are attached to the substrate. A thick copper shield placed under the coils prevents eddy currents in an underlying semiconductor, giving low loss even when fabricated on a moderately doped silicon substrate. For example, an inductor with three turns, each 200 μm wide and 535 μm in diameter, has an inductance of 4 nH, a high Q of 65 at 1 GHz, and a self-resonance frequency of 4.2 GHz. Using a thin aluminum shield made with standard integrated circuit processing instead of thick copper lowers the Q by about 25% due to a small amount of substrate coupling. By fabricating the inductor on a glass substrate without a shield, the self-resonance frequency and the effective inductance both rise because there is less parasitic capacitance, while Q is about the same. Taking the resistance due to the skin effect into account, this Q is close to the theoretical maximum possible to due to series resistance alone. While raising the coils off of the substrate improves performance, their height may be a limitation in applications where space is a constraint. The coils have been demonstrated to be stronger than 25-μm-diameter bond wires, which is sufficiently robust for use in plastic injection-molded packages.

The PARC solenoid process uses all low-temperature steps, enabling its use on wafers already containing circuitry. Fabrication begins with deposition of up to 7 μm of copper onto the wafer for the ground plane [14]. Approximately 12 to 15 μm of benzocyclobutene (BCB), a low-loss dielectric, are spun on to raise the coil up off of the substrate. Vias are opened in the BCB for the coil anchors and electrical contact to ground. A proprietary conductive sacrificial layer is sputtered on, followed by gold, a thicker layer of molybdenum-chromium (MoCr) alloy, and a gold passivation layer, for a 1.5-μm-thick metal stack (Figure 7.6). By increasing the pressure part way through the deposition, the stress of the MoCr film is more compressive on the bottom than on the top—an example of stress engineering. The metal stack is patterned and etched to form the shapes of flattened coil half-turns, with arrays of small etch holes in them. The photoresist is left on the metal stack as a selective etchant removes the sacrificial layer. Due to the stress gradient in the



(a) (b)

Figure 7.5 The PARC inductor: (a) scanning-electron micrograph (SEM) of a five-turn solenoid inductor (the locations of the sides of the turns before release are visible); and (b) SEM close up of the tops of the turns where the metal from each side meets, showing the interlocked ends. The etch holes have been filled with copper. (© 2003 IEEE [13].)

MoCr, the free ends curl upward, but do not reach their final destination due to the stiffness of the thick photoresist. When heated, the photoresist softens and gradually allows the free ends to bend further. By precisely controlling the stress gradient and the thickness of the MoCr film, the radius of curvature is such that the

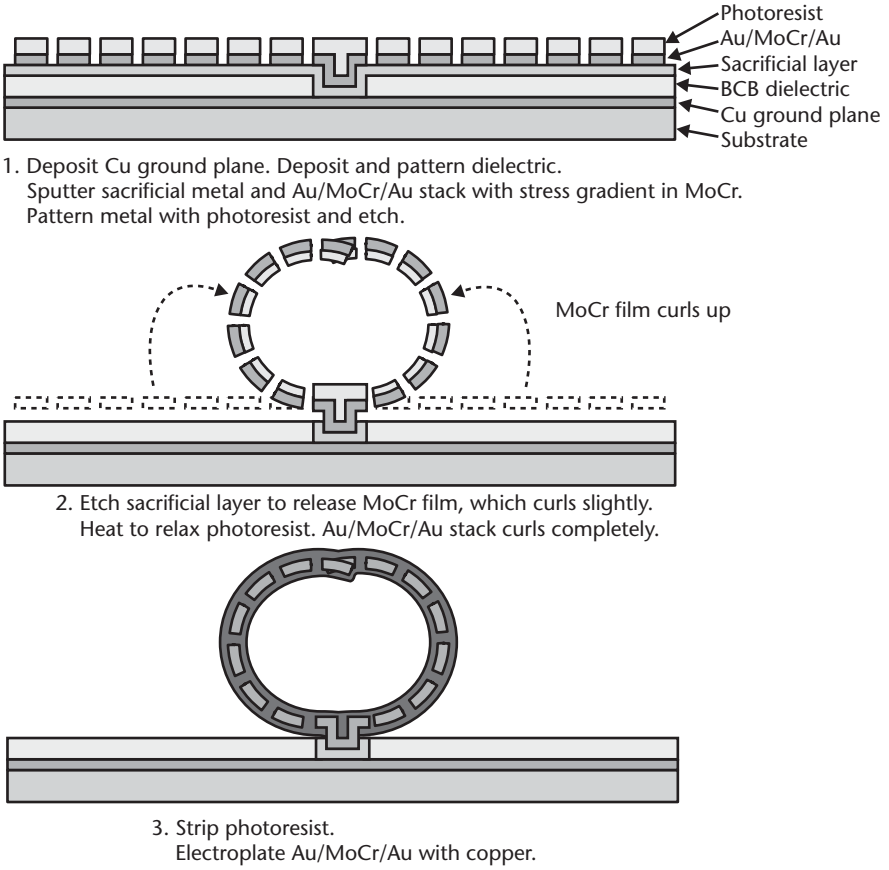


Figure 7.6 Illustration of the PARC inductor fabrication process.

two free ends meet. An interlocking tongue-in-groove structure aids in alignment and gives a small amount of process tolerance [Figure 7.5(b)]. The stack coils are electroplated with 5 to 8 μm of copper for a low resistance, using the gold on both sides as a seed layer. The copper also fills the etch holes and seals the seam where the two halves came together. Finally, the photoresist and any remaining release material are removed.

Microelectromechanical Resonators

A simple mechanical system of a spring with spring constant k and a mass m has a resonant frequency $f_r = (\frac{1}{2\pi})\sqrt{k/m}$ at which it naturally oscillates if the mass is moved and released (see Figure 7.7). If an external force drives the mass at this resonant frequency, the amplitude of the displacement rapidly grows until limited by losses in the system at steady state (the loss is known as *damping*). When driven at a frequency above or below the resonant frequency, the amplitude is smaller. In electronics, this is analogous to a series or parallel combination of capacitor and inductor, with a small series resistance.

As discussed earlier, the quality factor, Q, of a resonant electrical circuit or mechanical device is defined as the ratio of the maximum energy stored during a cycle to the energy lost per cycle. Thus, circuits or devices with higher Q values will have larger response (e.g., displacement) when driven at the resonant frequency (see Figure 7.8). Such circuits or devices also have a higher response peak and a narrower

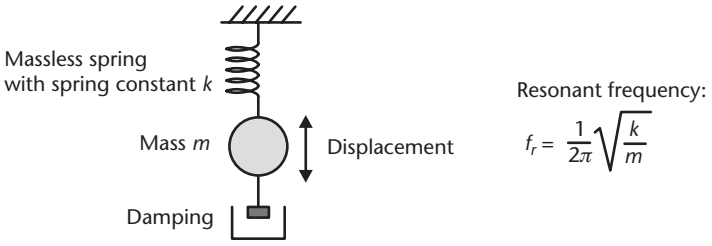


Figure 7.7 Illustration of a mechanical oscillator consisting of a spring, a mass, and a damping element that represents mechanical losses. When driven at its natural (resonant) frequency, the amplitude of the oscillation is greatest; at lower and higher frequencies, the amplitude is smaller.

$$Q = \frac{\text{maximum energy stored during cycle}}{\text{energy lost per cycle}} = \frac{\text{resonant frequency}}{\text{bandwidth at } 1/\sqrt{2} \text{ of maximum}}$$

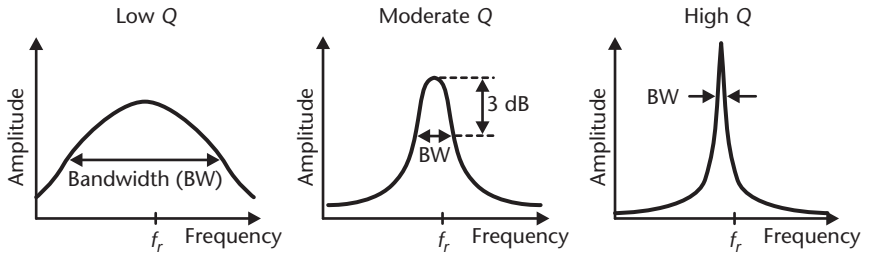


Figure 7.8 Illustration of the effect of the quality factor, Q, on the relationship between amplitude of oscillation and frequency.

bandwidth, which is the distance in frequency between the points of response that are $1/\sqrt{2}(-3\text{dB})$ below the response maximum. Bandwidth is also given a percentage of the center frequency.

Quartz crystals are presently at the core of every electrical resonant circuit because, historically, integrated electronic oscillators have not been able to achieve the large quality factors necessary for the stable operation of frequency-selective communications systems. A typical quartz crystal has a Q that reaches 10,000 or even higher. By comparison, the quality factor of an electrical filter consisting of a network of inductors, capacitors, and resistors (RLC network) is typically far less than 1,000, limited by parasitic resistive losses in the circuit. The quality factor has also an effect on insertion loss. For example, a simple bandpass filter consisting of a series inductor and capacitor, with parasitic resistance, in series with an output resistor, which has a center frequency at 16 MHz, a Q of 100, and a bandwidth of 2.9 MHz (18% of the center frequency) has an insertion loss of 0.8 dB—in other words, the signal suffers an undesirable attenuation of about 9%. The insertion loss increases further as the Q decreases. Quality factors above 1,000 are generally considered high for many electronic and RF applications. If micromechanical resonators can demonstrate high Q over a wide range of tunable frequencies, then integrating them with electronics will consequently lead to system miniaturization. The frequencies of interest cover the range between 800 MHz and 2.5 GHz for front-end wireless reception, as well as the intermediate frequencies[1] at 455 kHz and above.

Based on the equation for resonant frequency, it follows immediately that a reduction in size, which brings about a decrease in mass and stiffening of the spring, increases the resonant frequency. This is the basic argument for the micromachining of resonators. The various designs differ in their implementation of excitation and sense mechanisms.

Comb-Drive Resonators

One of the earliest surface-micromachined resonator designs [15], which is now commonly used in various MEM devices, is the interdigitated-finger comb-drive structure developed at the University of California, Berkeley, California (see Figure 7.9). This structure is comprised of folded springs supporting a *shuttle plate* that oscillates back and forth in the plane of the wafer surface. The folded springs relieve residual stress and give a more compact layout. An applied voltage, either positive or negative, generates an electrostatic force between the left anchor comb and shuttle comb that pulls the shuttle plate to the left in Figure 7.9. This electrical force F_e is given by $\frac{1}{2}(dC/dx)\,V^2$, where V is the applied voltage, and dC/dx is the rate of increase in capacitance as the finger overlap increases and is constant for a given design. Because the voltage is squared, the force is always attractive. When a sinusoidal ac voltage $v_a\cos(\omega t)$ is applied, where v_a is the amplitude and ω is the

1. A receiver converts the frequency of a selected incoming RF signal to a fixed intermediate frequency by heterodyning the signal with the local oscillator. This allows the remaining circuits in the receiver to remain precisely tuned to the intermediate frequency regardless of the frequency of the incoming signal. The following frequencies are generally considered intermediate frequency: 50 kHz, 100 kHz, 262 kHz, 455 kHz, 500 kHz, 9 MHz, 10.7 MHz, 45 MHz, and 75 MHz.

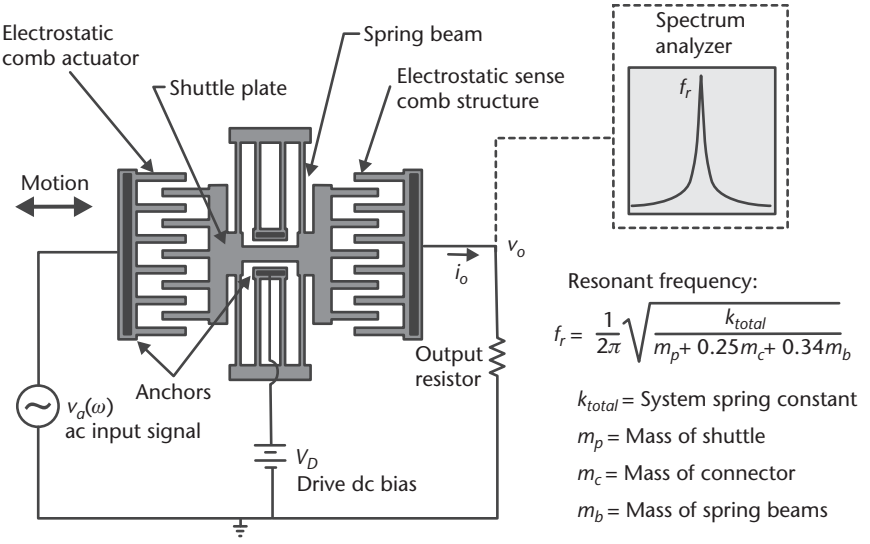


Figure 7.9 Illustration of a micromachined folded-beam comb-drive resonator. The left comb drive actuates the device at a variable frequency ω. The right capacitive-sense-comb structure measures the corresponding displacement by turning the varying capacitance into a current, which generates a voltage across the output resistor. There is a peak in displacement, current, and output voltage at the resonant frequency.

frequency in rad/s ($\omega = 2\pi f$), the force is proportional to $v_a^2 \times \cos^2(\omega t) = v_a^2 \times \frac{1}{2}[1 + \cos(2\omega t)]$. Thus, the force driving the resonator appears at a frequency of twice the input frequency, in addition to a dc component.

A frequency response different than the input frequency is not particularly useful for filters. To make a useful linear filter, a dc bias is superimposed so that the input across the comb is $V_D + v_a\cos(\omega t)$. The force is then proportional to $[V_D + v_a\cos(\omega t)]^2 = V_D^2 + 2V_D v_a\cos(\omega t) + v_a^2\cos^2(\omega t)$. In a linear filter, V_D is intentionally made much greater than v_a, so that the last term is negligible and the dominant time-varying drive force is at the input frequency ω. The final part of the filter is an output resistor attached to the sense comb on the right side of the structure in Figure 7.9. An output current $i_o = d(CV)/dt = V_D \bullet dC/dt = V_D(dC/dx) \bullet (dx/dt) = V_D(dC/dx) \bullet \omega \bullet x_{max}\sin(\omega t)$ flows through the output resistor, where x_{max} is the maximum displacement. Because x_{max} has a peak at the resonant frequency ω_r ($= 2\pi f_r$), the output current also peaks at ω_r.

Because this device only responds to a narrow range of frequencies, it can be used to set the frequency in a frequency-reference circuit [16]. It can also be used as a mixer in a heterodyne unit. Driving an anchor with a drive signal at frequency ω_d and the shuttle plate at a carrier frequency ω_c with a dc offset generates an electrostatic time-varying force that has a spectral signature at the fundamental frequencies ω_d and ω_c, at the sum and difference frequencies ($\omega_d + \omega_c$) and ($\omega_d - \omega_c$), and at the second harmonics, $2\omega_d$ and $2\omega_c$, as discussed earlier. Only the frequency near the mechanical resonance is passed to the output.

The spring constant k of a single clamped-clamped beam bending to the side is given by $k_{beam} = E \bullet t \bullet (w/L)^3$, where E is the Young's modulus, t is the beam thickness, w is the width, and L is the length. For the structure shown in Figure 7.9, the

total spring constant for the system of spring beams is $k_{total} = 2\ k_{beam}$. Because the springs do not move as much as the main shuttle mass, only a fraction of their mass is added to that of the shuttle mass in determining the resonant frequency.

A representative comb-drive resonator made of polycrystalline silicon using standard surface-micromachining techniques has beams with a thickness 2 μm, widths of 2 μm, and lengths of 185 μm, resulting in a system spring constant of 0.65 N/m. With an effective motional mass equal to 5.7×10^{-11} kg, the structure resonates at 17 kHz [17]. Keeping the same beam thickness and width but reducing the length to 33 μm gives a structure that resonates at 300 kHz [18]. The Q can be over 50,000 in vacuum but rapidly decreases to below 50 at atmospheric pressure due to viscous damping in air [19]. Thus, vacuum packaging is necessary to commercialize these high-Q devices.

To attain a higher resonant frequency, the total spring constant must be increased or the motional mass must be decreased. The former is done by increasing the beam width and decreasing its length; the latter is difficult to do while retaining a rigid shuttle with the same number of comb fingers. Using electron-beam lithography to write submicron linewidths, single-crystal silicon beams with lengths of 10 μm and widths of 0.2 μm reached a resonant frequency of 14 MHz [20]. Alternatively, while using the same resonator dimensions, the resonant frequency can be increased by using a material with a larger ratio of Young's modulus, E, to density, ρ, than silicon. Metals known in engineering for their high stiffness-to-mass ratio, such as aluminum and titanium, have a ratio E/ρ that is actually lower than for silicon. Two materials with higher E/ρ ratios are silicon carbide and polycrystalline diamond; the latter is a research topic for high-frequency resonators.

Beam Resonators

To build a micromachined structure with higher resonant frequency than that readily achievable with a comb drive, the mass must be further reduced. Beam resonators have been studied extensively at the University of Michigan, Ann Arbor [21–23], for this purpose, and Discera, Inc., of Ann Arbor, Michigan, is commercializing them for reference frequency oscillators to replace quartz crystals in cellular phones. The advantages include a much smaller size, the ability to build several different frequency references on a single chip, higher resonant frequencies, more linear frequency variation with temperature over a wide range, and the ability to integrate circuitry, either on the same chip or on a circuit chip bonded to the MEM chip, all at a lower cost than the traditional technology.

The simplest beam resonator is rigidly clamped on both ends and driven by an underlying electrode (see Figure 7.10). A dc voltage applied between the beam and the drive electrode causes the center of the beam to deflect downward; removal allows it to travel back upward. An ac drive signal v_a causes the beam to flex up and down. As with the comb-drive actuator, when a large dc bias V_D is superimposed to the ac drive signal, the beam oscillates at the same frequency as the drive signal. At resonance, the deflection amplitude is at its greatest. An example polysilicon beam is 41 μm long, 8 μm wide, 1.9 μm thick, with a gap of 130 nm [21]. Applying voltages V_D = 10V and v_a = 3 mV, the measured resonant frequency is 8.5 MHz, the quality factor Q is 8,000 at a pressure of 9 Pa, and the deflection amplitude is a mere 4.9 nm at the center of the beam. At atmospheric pressure, Q drops to less than

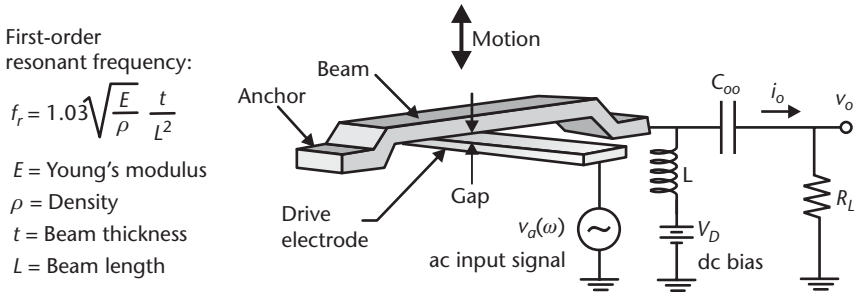


Figure 7.10 Illustration of a beam resonator and a typical circuit to measure the signal. The beam is clamped on both ends by anchors to the substrate. The capacitance between the resonant beam and the drive electrode varies with the deflection.

1,000 due to viscous damping, and the deflection at resonance falls by a corresponding ratio. To maintain a high Q in its product, Discera uses an on-chip, vacuum-sealed cap over the resonators.

The dc bias also adds a downward electrostatic force. This force varies with distance and opposes the mechanical restoring force of the beam, making the effective mechanical spring constant of the system smaller. The resonant frequency falls by a factor proportional to $\sqrt{1-\left(CV_D^{\,2}/k^2g^2\right)}$, where C is the initial capacitance, k is the mechanical spring constant, and g is the gap without a dc bias. Thus, the resonant frequency can be electrically tuned.

In contrast to the two-port resonators described later, this single-beam resonator is a one-port device with only one pair of external leads. The resonator appears as a time-varying capacitor, $C(\omega)$, because the capacitance between the beam and the drive electrode changes with the deflection. A simple electrical circuit using external passive components is necessary to measure an electrical signal from the resonator [22]. The circuit includes a shunt blocking inductor, L, and a series blocking capacitor, C_∞ (see Figure 7.10). With a large dc bias V_D, the dominant output current at the input frequency ω is $i_o = V_D dC/dt$. At high frequency, the inductor L is an open circuit, and the output capacitor C_∞ is a short, so that i_o flows through the load resistor R_L. In practice, the load resistance may be the input impedance of the measurement equipment. Alternatively, a transimpedance amplifier, which amplifies an input current and outputs a voltage, can substitute the load resistance.

One of the key requirements of a frequency reference is stability over the operating temperature range. As the temperature rises, the Young's modulus for most materials falls, resulting in a lower spring constant and therefore a lower resonant frequency. For polysilicon clamped-clamped beams, the rate at which the resonant frequency falls is -17×10^{-6}/K (ppm/K), compared to the -1×10^{-6}/K range for AT-cut quartz crystals (the exact value varies due to production variation) [23]. A solution to this problem, being implemented in Discera products, is variable electrical stiffness compensation.

When an electrode with a dc bias V_C is placed over a beam, an effective second electrical spring is added to the system, which also reduces the overall system spring constant [see Figure 7.11(a)]. By mounting the ends of this top electrode on top of metal supports with a faster thermal expansion rate than that of the polysilicon electrode, the gap increases with temperature. This reduces the electrical spring constant

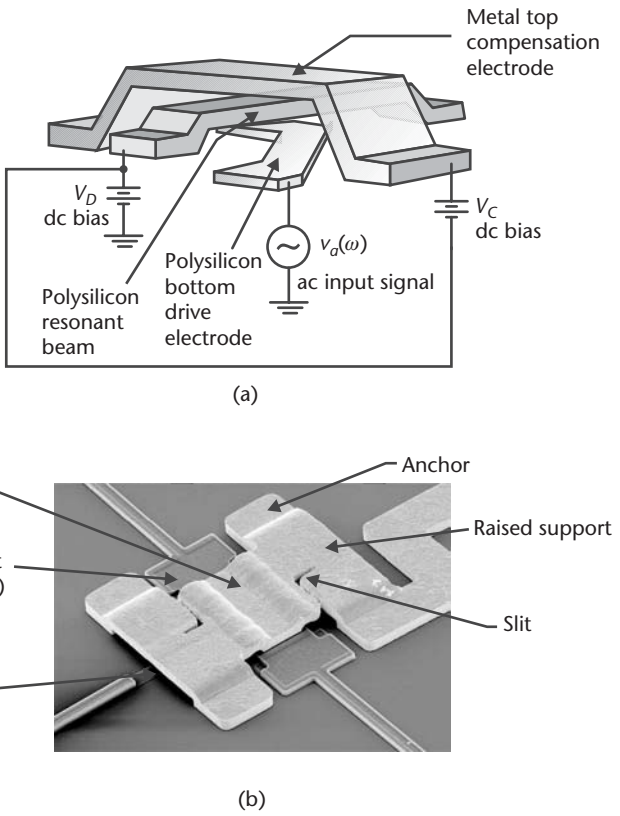


Figure 7.11 Illustration of the compensation scheme to reduce sensitivity in a resonant structure to temperature. A voltage applied to a top metal electrode modifies through electrostatic attraction the effective spring constant of the resonant beam. Temperature changes cause the metal electrode to move relative to the polysilicon resonant beam, thus changing the gap between the two layers. This reduces the electrically induced spring constant opposing the mechanical spring while the mechanical spring constant itself is falling, resulting in their combination varying much less with temperature. (a) Perspective view of the structure [23], and (b) scanning electron micrograph of the device. (*Courtesy of:* Discera, Inc., of Ann Arbor, Michigan.)

opposing the mechanical spring, while the mechanical spring constant itself is falling, resulting in their combination varying much less with temperature (down to $+0.6 \times 10^{-6}$/K in prototypes [23]).

For process compatibility the entire top electrode is made of metal, which also expands faster laterally than the underlying silicon substrate. Because it is clamped at the ends, it undesirably bows upward unless measures are taken to prevent this. By suspending the ends of this beam off of the substrate and putting slits near its ends [see Figure 7.11(b)], bowing is greatly reduced, from 6 nm down to 1 nm when heated to 100°C. When appropriately biased, this reduces the frequency shift with temperature to only -0.24×10^{-6}/K, comparable to the best quartz crystals [23]. Design specifications for this prototype beam are a length of 40 μm, width of 8 μm, thickness of 2 μm, gap below the resonant beam during operation of 50 nm, and gap above the beam of about 250 nm. With a beam-lower electrode dc bias V_D of 8V and a beam-upper electrode voltage V_C also of 8V, the resonant frequency is 9.9 MHz with a Q of 4,100. For the Discera products to be used as cellular phone

reference oscillators, beams are designed for resonant frequencies including 19.2 MHz and 76.8 MHz for code-division multiple access (CDMA) wireless networks and 26 MHz for Global System for Mobile Communications (GSM) networks.

The bottom electrode and the resonant beam are fabricated from polysilicon using standard surface micromachining steps, with a sacrificial silicon dioxide layer in between [23]. A sacrificial oxide layer is also formed on top of the resonant beam, followed by a sacrificial nickel spacer on the sides of the beam. Gold electroplated through a photoresist mask forms the top metal electrode. Finally, the nickel and silicon dioxide are etched away to leave the freestanding beams.

Coupled-Resonator Bandpass Filters

The resonators just described have a very narrow bandpass characteristic, making them suitable for setting the frequency in an oscillator circuit but not for a more general bandpass filter. Bandpass filters pass a range of frequencies, with steep roll-off on both sides. Two or more microresonators, of either the comb-drive or clamped-clamped beam type, can be linked together by weak springs or flexures to create useful bandpass filters (see Figure 7.12).

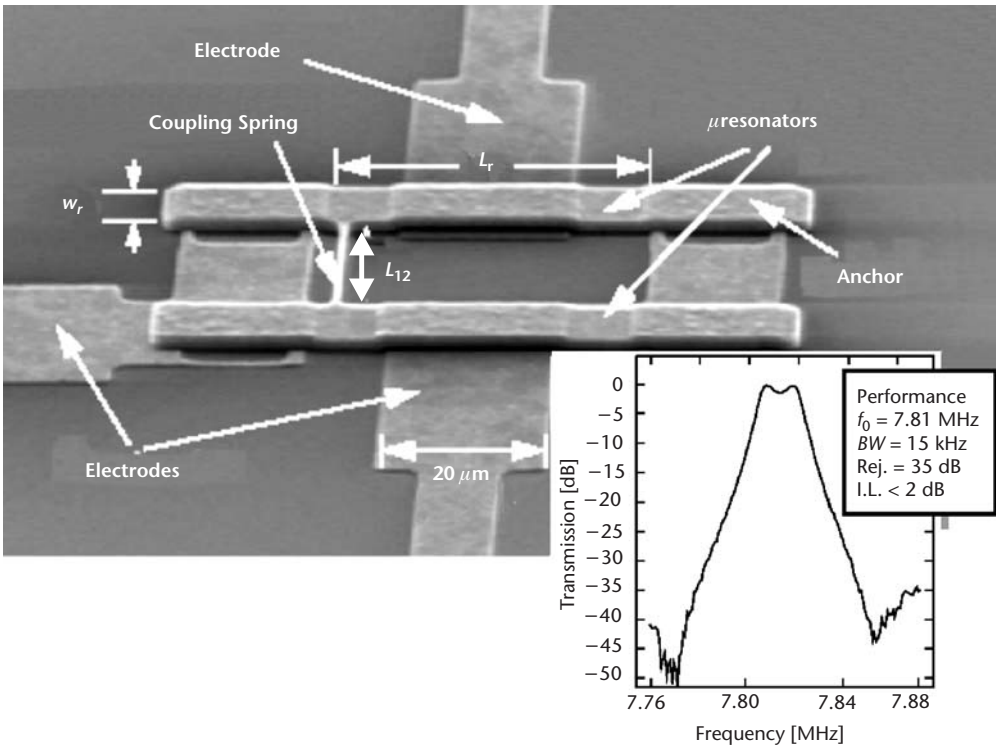


Figure 7.12 Scanning electron micrograph of a polysilicon surface micromachined bandpass filter consisting of two clamped resonant beams coupled by a weak intermediate flexure spring. The excitation and sensing occur between the beams and electrodes beneath them on the surface of the substrate. Each resonant beam is 41 μm long, 8 μm wide, and 2 μm thick. The coupling flexure is 20 μm long and 0.75 μm wide. (© 1998 IEEE [24].)

To visualize this complex effect, let us imagine two physically separate but identical simple resonators consisting of a mass and a spring. These resonators can freely oscillate at the natural frequency determined by the mass and the spring constant. Adding a weak and compliant flexure or spring between the two masses (see Figure 7.13) restricts the allowed oscillations of this two-body system. The two masses can move either in phase or out of phase with respect to each other; these are the two *oscillation modes* of the system. When the motions are in phase, there is no relative displacement between the two masses and, consequently, no restoring force from the weak flexure. The oscillation frequency of this first mode is then equal to the natural frequency of a single resonator. When the two masses move out of phase with respect to each other, however, their displacements are in opposite directions at any instant of time. This motion produces the largest relative displacement across the coupling flexure, thereby resulting in a restoring force, which, according to Newton's second law, provides a higher oscillation frequency. The physical coupling of the two masses effectively split the two overlapping resonant frequencies (of the two identical resonators) into two distinct frequencies, with a frequency separation dependent on the stiffness of the coupling flexure. In physics, it is said that the coupling lifts the degeneracy of the oscillation modes. For a very compliant coupling spring, the two split frequencies are sufficiently close to each other that they effectively form a narrow passband. Increasing the number of coupled oscillators in a

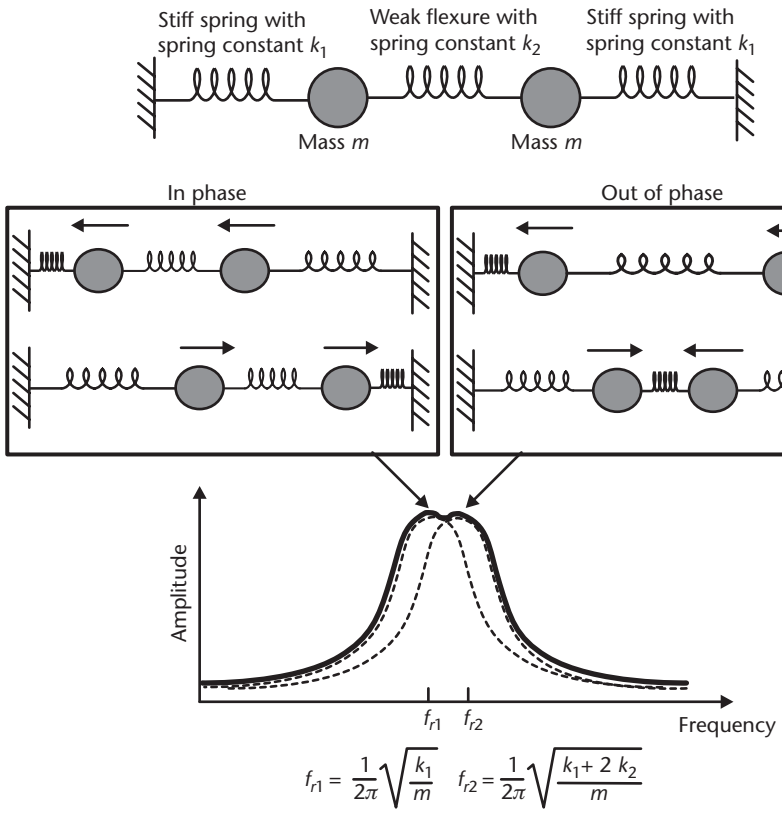


Figure 7.13 Illustration of two identical resonators, each with a mass and spring, coupled by a weak and compliant intermediate flexure. The system has two resonant oscillation modes, for in-phase and out-of-phase motion, resulting in a bandpass characteristic.

linear chain widens the extent of this passband but also increases the number of ripples. In general, the total number of oscillation modes is equal to the number of coupled oscillators in the chain.

Coupled-resonator filters are two-port devices, with a two-lead input and two-lead output. An ac voltage input drives the filter, while the output is taken in the same method as that for a single resonator: a dc bias is applied. The current due to the capacitance change, $V_D\ dC/dt$, is the output, which is typically fed to a transimpedance amplifier to generate an output voltage. From the perspective of an electrical engineer, a dual electrical network models the behavior of a filter made of coupled micromechanical resonators. The dual of a spring-mass system is a network of inductors and capacitors (LC network): The inductor is the dual of the mass (on the basis of kinetic energy), and the capacitor is the dual of the spring (on the basis of potential energy). A linear chain of coupled undamped micromechanical resonators becomes equivalent to an LC ladder network. This duality allows the implementation of filters of various types using polynomial synthesis techniques, including Butterworth and Chebyshev common in electrical filter design. Widely available "cookbooks" of electrical filters provide appropriate polynomial coefficients and corresponding values of circuit elements [18].

Film Bulk Acoustic Resonators

Another method of creating microelectromechanical bandpass filter is to use a piezoelectric material. By sandwiching a sheet of piezoelectric material with a reasonably high d_{33} (see Chapter 3) and low mechanical energy loss between two electrodes, a resonator is created [see Figure 7.14(a)]. When an ac signal is applied across the piezoelectric, an acoustic wave, traveling at the speed of sound in the material, is generated. If the top and bottom surfaces of the device are in air or vacuum, there is an acoustic impedance mismatch, and the wave is reflected back and forth through the thickness. When the acoustic wavelength is equal to twice the thickness, a standing wave is formed (mechanical resonance) and the electrical impedance is low [see Figure 7.14(b)]. The frequency response of such devices is commonly modeled by the simplified L-C-R electrical network shown in Figure 7.14(c). The series inductance and capacitance in the model represent the kinetic energy of the moving mass and the stored energy due to compression and expansion of the material, respectively, while the series resistor represents energy loss. This resistance is relatively small with a good design and process, enabling quality factors of over 1,000 in production devices. There is also a significant electrical capacitance between the plates, represented by the parallel capacitor. The series capacitor and inductor in this system have a series resonance—the low impedance in Figure 7.14(b). Due to the parallel capacitor, the system also predicts a separate, parallel resonance—the high impedance in Figure 7.14(b).

The goal of a bandpass filter, such as those linking the input or output circuitry to the antenna of a cellular phone, is to transmit a narrow range of frequencies with low loss and filter out both higher and lower frequencies. To make a bandpass filter, FBARs are placed in a ladder network such as that shown in Figure 7.14(d) [25]. The series FBARs are designed to have the same series-resonant frequency and corresponding low impedance, which transmits the desired frequency with low loss [see Figure 7.14(e)]. These devices do not transmit higher frequencies due to the high

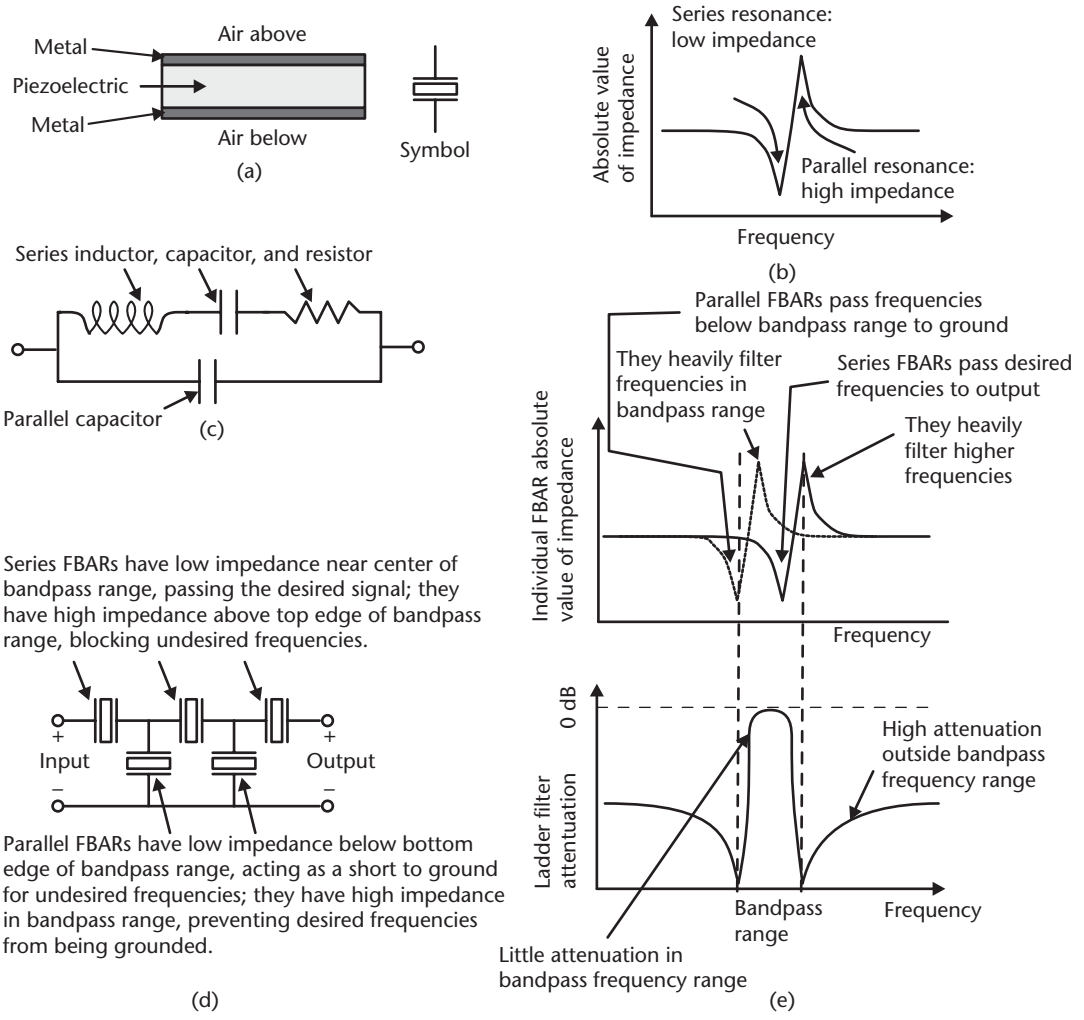


Figure 7.14 Film bulk acoustic resonator (FBAR): (a) cross section of an FBAR and symbol; (b) impedance versus frequency of an individual FBAR; (c) equivalent electrical circuit; (d) FBARs in ladder filter; and (e) impedance versus frequency for two FBARs and relation to attenuation versus frequency of ladder filter.

impedance resulting from the parallel resonance just above the series-resonant frequency. The parallel FBARs are designed to have a lower series-resonant frequency, shorting undesired signals to ground but not affecting the desired frequency. The result is a transmission curve such as that shown in Figure 7.14(e) for a commercial device [26]. Adding more stages provides more filtering of undesired frequencies—but at the cost of more attenuation of the desired frequencies.

Agilent Technologies, Inc., of Palo Alto, California, started marketing FBAR-based RF bandpass filters for cellular phone handsets in 2001. There is a great consumer demand for smaller cellular phones, and FBAR filters are one of the microelectromechanical devices that have helped to meet this demand by being much smaller than the ceramic surface-acoustic wave devices they replaced. They also enable new filter applications by meeting very sharp frequency filter roll-off specifications and being able to handle power of over 1W [25]. In most applications

to date, pairs of FBAR filters have been placed at the antenna of a cellular phone to form a duplexer. One bandpass filter allows signal transmission from the output power amplifier to the antenna (e.g., 1.85–1.91 GHz for PCS); the other has a different bandpass that transmits received signals from the antenna to the input low-noise amplifier (e.g., 1.93–1.99 GHz for PCS).

The fabrication of an Agilent FBAR filter, the exact details of which are proprietary, begins with a high-resistivity silicon wafer (see Figure 7.15) [27, 28]. The high resistivity is needed to reduce losses due to eddy currents; alternatively, an insulating substrate such as glass could be used. A cavity a few micrometers deep is etched into the silicon. The silicon is thermally oxidized for electrical isolation. Phosphosilicate glass (PSG) is deposited using LPCVD sufficiently thick to fill the cavity. Chemical-mechanical polishing then removes all of the PSG outside of the cavity and planarizes the wafer surface. At this point, the PSG in the cavity has a very smooth surface, which is critical for the later deposition of aluminum nitride (AlN). About 0.1 μm of molybdenum (Mo) is sputtered and patterned to form the lower electrode. Molybdenum is chosen for its suitably low electrical resistivity, low mechanical loss, and process compatibility. This is followed by sputtering and patterning of the aluminum nitride piezoelectric layer. The AlN thickness is chosen so that the thickness of the stack is one half of a wavelength of the speed of sound in these materials for the desired resonant frequency. For example, if the speed of sound through the thickness of the AlN is 11.4 km/s and the desired tuning frequency is 1.88 GHz, then the wavelength is 6.1 μm and the thickness is chosen to be approximately 3 μm (the electrodes add acoustic additional path length, so the piezoelectric layer is slightly thinner than this). Thickness control is critical, as it sets the resonant frequency. Next, another layer of molybdenum is sputtered, followed by gold, which adheres to the Mo. The gold is patterned for the bond pads, followed by patterning of the Mo for the top electrode. A small area of extra metal is added on top of some resonators to mass-load them and reduce their acoustic resonant frequency for the ladder filter [28]. Alternatively, a small amount of metal could be removed from the surface to

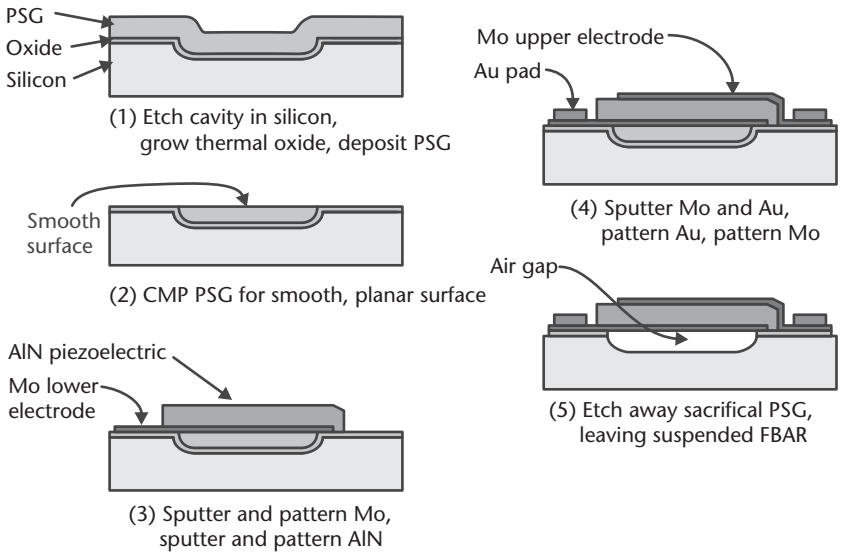


Figure 7.15 Illustration of an example FBAR fabrication process.

raise the acoustic resonant frequency. Finally, the PSG under the resonator is etched away in a hydrofluoric acid solution, leaving the freestanding structure. Access for the etching acid is provided through holes at the edges of the resonator structure. Releasing the structure has alternatively been performed by orientation-dependent etching of the silicon from the back side of the wafer, but this requires double-sided alignment and starting the back side etch with a window much larger than the area of the resonator, greatly increasing the chip area and cost [28].

When FBARs in a filter block applied signals, some of the power is absorbed (the rest is passed through or reflected). Each FBAR's area must be large enough to dissipate heat without causing a problem, such as a significant shift in the frequency characteristic. One issue with such filters with a mechanical resonance back and forth through the thickness of the piezoelectric material is that some acoustic energy is unintentionally coupled into the plane of the material. To prevent standing waves from forming in plane, the walls of the Agilent FBAR are not parallel but rather at an angle to each other; implementations include forming a nonparallelogram quadrilateral or an irregular pentagon [29].

Microelectromechanical Switches

Microelectromechanical switches have many potential applications in electronics. In cellular phones, they can rapidly isolate and connect the send and receive channels to a common antenna, as well as performing less frequent reconfigurations for different communications standards (e.g., global system for mobile communications, code-division multiple access, or time-division multiple access). At extremely high frequencies (above 30 GHz), the wavelength becomes sufficiently short that small phased arrays of antennas can be fabricated for radar applications: banks of switches can rapidly reconfigure phase shifters to drive the phased arrays to orient the transmitted signal in different directions. Automated test and measurement equipment also use arrays of switches for the application of power and signals to devices under test.

The key desirable parameters in RF switches are low insertion loss and return loss (reflection) in the closed state, high isolation in the open state, high linearity (typically quoted as level of third-order harmonic), high power-handling capability during switching, low operating voltage (for portables), high reliability (particularly a large number of cycles before failure), small size, and low cost. As will be seen, there are tradeoffs among various combinations of these parameters. For microelectromechanical switches to be designed into new products, they must surpass the performance of, or offer some other advantage such as small size or cost over, existing switch technologies such as gallium arsenide FETs, silicon *p-i-n* diodes, and traditional electromagnetic relays.

Many different micromechanical RF switch prototypes have been fabricated, drawing on the various actuation techniques discussed in Chapter 4. The most common are electrostatically driven, which is appealing for handheld and satellite applications for its negligible power consumption when holding. Most electrostatically driven switches have a membrane or a cantilever containing one contact, which is suspended over the substrate supporting another contact (see Figures 7.16 and 7.17).

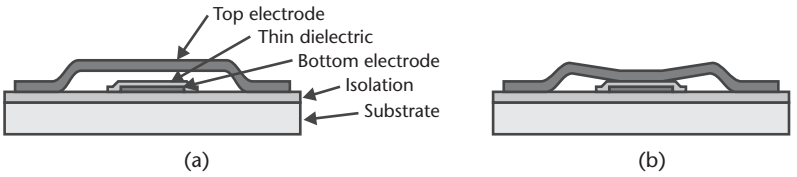


Figure 7.16 Illustration of a membrane switch: (a) In the open state, the metal lines act as a waveguide, with the sides being ground and the signal propagating down the center line. (b) In the closed state, application of a dc voltage pulls the top ground membrane down to short the signal line. If there is a thin dielectric as shown, the impedance is low only at high frequency. (*After:* [30].)

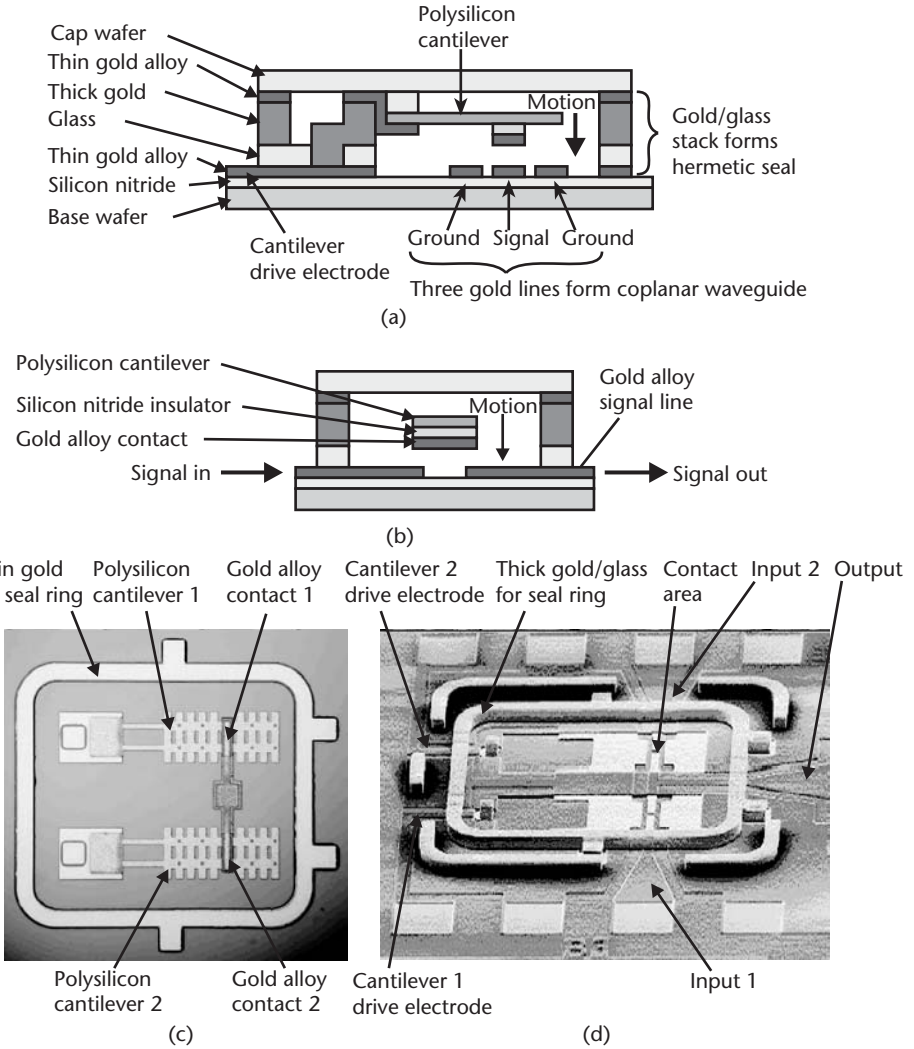


Figure 7.17 Illustration of the MicroAssembly cantilever switch: (a) Cross section along the length of the cantilever, showing the coplanar waveguide. (b) Cross section across the width of the cantilever, showing the signal contact region. (c) Micrograph of the top wafer of a single-pole, double-throw switch, containing the cantilever, before assembly. (d) Scanning electron micrograph of the bottom wafer, containing the coplanar waveguide and seal ring, before assembly. (*Courtesy of:* MicroAssembly Technologies of Richmond, California.)

Membrane Shunt Switch

In a membrane-switch implementation from the University of Michigan, Ann Arbor, Michigan, a 2-μm-thick layer of gold is suspended 2 μm above a 0.8-μm-thick gold signal line, which is coated with about 0.15 μm of insulating silicon nitride [30]. The membranes have a span of 300 μm and lengths of 20 to 140 μm. Application of a 15-V dc voltage to the signal line (in addition to the ac signal) pulls the gold membrane down to the nitride, shunting the signal line to ground [31]. The use of an insulator prevents this switch from working at dc and low frequency but is expected to be more reliable than metal-to-metal contacts. The sides of the gold membrane are supported by wide strips of the same gold film, which, with the signal line, form the coplanar waveguide needed for microwave signals.

In the closed state, the connection is made by capacitive coupling, which is only useful at high frequency. Insertion loss in the closed state is less than 0.6 dB over the range of 10–40 GHz, with a return loss of less than –20 dB. The silicon nitride could be made even thinner for a lower closed-state capacitance and lower insertion loss, but it is already at the minimum thickness required to prevent breakdown with the required dc operation voltage. In the open state, there is clearly a capacitor that causes undesired coupling. The gap could be increased for greater isolation (at least 20 dB is desired), but this would require an even greater actuation voltage. The measured ratio of closed to open capacitance in this design is about 17. The loss in the up state is the same as for the coplanar waveguide alone.

Cantilever Series Switch

Several cantilever-type switches with metal-to-metal contacts (also known as relays) are under commercial development by companies such as Teravicta Technologies of Austin, Texas; Radant MEMS, Inc., of Stowe, Massachusetts; and MicroAssembly Technologies, Inc., of Richmond, California (see Figure 7.17). In the switch from MicroAssembly Technologies, a voltage applied to a cantilever creates a potential between the free end and underlying ground lines. The cantilever is pulled down, with the structure stopping when a separate gold-alloy contact on the cantilever mates with contacts below, bridging the gap between the contacts. Coplanar waveguides maintain microwave-signal integrity. A single pole, double throw switch (two inputs and one output) is shown in the micrographs in Figure 7.17; switches have been designed with one up to four inputs.

The area of greatest concern in switches of any size is reliability of the contact itself. In the MicroAssembly switch, a proprietary gold alloy is employed. Special attention is paid to the mechanics of the switch closure. For example, a relatively high closing force aids in keeping the electrical resistance low over the life of the switch. Melting and sputtering of the metal to create a self-renewing contact is being studied. Cleanliness is also critical: the contacts can be contaminated by volatile organic compounds far below the part-per-billion level. A well-controlled switch environment is established by the use of integrated hermetic packaging (see Chapter 8). As seen in Figure 7.17, a multilayer gold/glass ring surrounds the switch, while allowing signals to flow in and out. Another substrate forms the cap, which is bonded at the wafer level for low handling cost. The use of integrated packaging gives this device a level of completeness not found in many prototypes and produces

a smaller size and lower bill-of-materials cost than if a separate hermetic package were used.

The gap between the cantilever and the underlying electrodes must be sufficiently large that the isolation is high when open. Furthermore, the cantilever must be stiff enough that it is not damaged and closure does not accidentally occur when the device is shocked (switches have demonstrated a shock tolerance of 30,000G). These criteria lead to a higher actuation voltage than is available in many systems. To resolve this problem, charge-pump circuitry supplies the needed drive voltage from a 1.5-V input. When closed, which takes 10 μs, the measured insertion loss of the switch is less than 0.2 dB from dc to 2 GHz. For the package alone, the insertion loss is nearly "invisible" to the circuit at 0.06 dB, which has enabled this packaging scheme to be used for other RF devices as well as switches. The isolation when open is 40 dB. Goals are an insertion loss of 0.2 dB over 24–40 GHz, a lifetime of 10^{11} cycles, and 1-W cold-switched power handling.

Summary

The most notable members of the RF MEMS family, micromachined variable capacitors, inductors, resonators, filters, and switches were described, with research or commercial examples of each. There are different advantages of these devices, compared to their conventional counterparts, including smaller size, lower cost, higher Q, lower loss, and the ability to be integrated on the same chip as circuitry. A common theme in RF MEMS is the reduction of parasites. We are presently at the dawn of an era of commercial use of RF MEMS.

References

[1] Ramo, S., J. R. Whinnery, and T. van Duzer, *Fields and Waves in Communications Electronics*, 3rd ed., New York: John Wiley, 1994.

[2] Solymar, L., and D. Walsh, *Lectures on the Electrical Properties of Materials*, 3rd ed., Oxford, United Kingdom: Oxford Univeristy Press, 1985.

[3] Nguyen, C. T. -C., L. P. B. Katehi, and G. M. Rebeiz, "Micromachined Devices for Wireless Communications," *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1756–1786.

[4] Young, D. J., et al., "A Low-Noise RF Voltage-Controlled Oscillator Using On-Chip High-Q Three-Dimensional Inductor and Micromachined Variable Capacitor," *Technical Digest of Solid-State Sensor and Actuator Workshop*, Hilton Head, NC, June 1998, pp. 128–131.

[5] Young, D. J., and B. E. Boser, "A Micromachined Variable Capacitor for Monolithic Low-Noise VCOs," *Technical Digest of Solid-State Sensor and Actuator Workshop*, Hilton Head, SC, June 1996, pp. 86–89.

[6] Dec, A., and K. Suyama, "Microwave MEMS-Based Voltage-Controlled Oscillators," *IEEE Transactions on Microwave Theory and Techniques,* Vol. 48, No. 11, November 2000, pp. 1943–1949.

[7] U.S. Patent 6,549,394, April 15, 2003.

[8] Dec, A., and K. Suyama, "Micromachined Electro-Mechanically Tunable Capacitors and Their Application to IC's," *IEEE Transactions on Microwave Theory and Techniques,* Vol. 46, No. 12, December 1998, pp. 2587–2596.

[9] Yao, J. J., S. Park, and J. DeNatale, "High Tuning-Ratio MEMS-Based Tunable Capacitors for RF Communications Applications," *Technical Digest of Solid-State Sensor and Actuator Workshop*, Hilton Head, NC, June 1998, pp. 124–127.

[10] Yao, J. J., et al., "A Low Power/Low Voltage Electrostatic Actuator for RF MEMS Applications," *Technical Digest of Solid-State Sensor and Actuator Workshop*, Hilton Head, NC, June 2000, pp. 246–249.

[11] Yao, J. J., "RF MEMS from a Device Perspective," *Journal of Micromechanics and Microengineering*, Vol. 10, 2000, pp. R9–R38.

[12] Ulrich, R., and L. Schaper, "Putting Passives in Their Place," *IEEE Spectrum*, Vol. 40, No. 7, July 2003, pp. 26–30.

[13] Van Schuylenbergh, K., et al., "Low-Noise Monolithic Oscillator with an Integrated Three-Dimensional Inductor," *Technical Digest of International Solid-State Circuits Conference*, San Francisco, CA, February 2003, pp. 392–393.

[14] Chua, C. L., et al., "Out-Of-Plane High-*Q* Inductors On Low-Resistance Silicon," *Journal of Microelectromechanical Systems*, Vol. 12, No. 6, December 2003, pp. 989–995.

[15] Tang, W. C., T. -C. H. Nguyen, and R. T. Howe, "Laterally Driven Polysilicon Resonant Microdevices," *Sensors and Actuators*, Vol. 20, 1989, pp. 25–32.

[16] Nguyen, C. T. -C., and R. T. Howe, "CMOS Micromechanical Resonator Oscillator," *Technical Digest of International Electron Devices Meeting*, Washington, D.C., December 1993, pp. 199–202.

[17] Nguyen, C. T. -C., "Frequency-Selective MEMS for Miniaturized Communications Devices," *Proceedings of 1998 IEEE Aerospace Conference*, Vol. 1, Snowmass, CO, March 1998, pp. 445–460.

[18] Wang, K., and C. T. -C. Nguyen, "High-Order Micromechanical Electronic Filters," *Technical Digest of IEEE 1997 International Micro Electro Mechanical System Workshop*, Nagoya, Japan, January 26–30, 1997, pp. 25–30.

[19] Nguyen, C. T. -C., and R. T. Howe, "Quality Factor Control for Micromechanical Resonantors," *Technical Digest of International Electron Devices Meeting*, San Francisco, CA, December 1992, pp. 505–508.

[20] McMillan, J. A., et al., "High Frequency Single Crystal Silicon Resonant Devices," *Proceedings of the 38th Annual Symposium and Topical Conference of the American Vacuum Society*, 1991.

[21] Bannon, F. D., J. R. Clark, and C. T. -C. Nguyen, "High-Q HF Microelectromechanical Filters," *IEEE Journal of Solid-State Circuits*, Vol. 35, No. 4, April 2000, pp. 512–526.

[22] Wang, K., A. -C. Wong, and C. T. -C. Nguyen, "VHF Free-Free Beam High-*Q* Micromechanical Resonators," *Journal of Microelectromechanical Systems*, Vol. 9, No. 3, September 2000, pp. 347–360.

[23] Hsu, W. -T., and C. T. -C. Nguyen, "Stiffness-Compensated Temperature-Insensitive Micromechanical Resonators," *Technical Digest of 15th IEEE Conference on Micro Electro Mechanical Systems*, Las Vegas, NV, January 2002, pp. 731–734.

[24] Nguyen, C. T. -C., "Frequency-Selective MEMS for Miniaturized Communications Devices," *Proc. 1998 IEEE Aerospace Conference*, Vol. 1, Snowmass, CO, March 21–28, 1998, pp. 445–460.

[25] Larson III, J. D., et al., "A BAW Antenna for the 1900 MHz PCS Band," *Proc. 1999 Ultrasonics Symposium*, pp. 887–890.

[26] "Agilent ACPF-7001 High Rejection Tx Filter for US PCS Band Data Sheet," Agilent Technologies, Inc., Palo Alto, CA, March 30, 2003.

[27] U. S. Patent 6,060,818, May 9, 2000.

[28] Ruby, R. C., et al., "Thin Film Bulk Wave Acoustic Resonators (FBAR) for Wireless Applications," *Proc. 2001 IEEE Ultrasonics Symposium*, pp. 813–821.

[29] U. S. Patent 6,215,375, April 10, 2001.

[30] Muldavin, J. B., and G. M. Rebeiz, "High-Isolation CPW MEMS Shunt Switches—Part 1: Modeling," *IEEE Transactions on Microwave Theory and Techniques*. Vol. 48, No. 6, June 2000, pp. 1045–1052.

[31] Muldavin, J. B., and G. M. Rebeiz, "High-Isolation CPW MEMS Shunt Switches—Part 2: Design," *IEEE Transactions on Microwave Theory and Techniques*. Vol. 48, No. 6, June 2000, pp. 1053–1056.

Selected Bibliography

De Los Santos, H. J., *RF MEMS Circuit Design for Wireless Communications*, Norwood, MA: Artech House, 2002.

Nguyen, C. T. -C., L. P. B. Katehi, and G. M. Rebeiz, "Micromachined Devices for Wireless Communications," *Proceedings of the IEEE*, Vol. 86, No. 8, August 1998, pp. 1756–1786.

Rebeiz, G. M., *RF MEMS: Theory, Design and Technology,* New York: Wiley, 2003.

Yao, J. J., "RF MEMS from a Device Perspective," *Journal of Micromechanics and Microengineering*, Vol. 10, 2000, pp. R9–R38.

CHAPTER 8

Packaging and Reliability Considerations for MEMS

> "Reality has surpassed fantasy. We're like kids in a candy store."
> —*Art Thompson, tactical activity lead of the NASA Mars Exploration Rovers mission after landing on Mars, January 2004.*

Packaging is the process, industry, and methods of "packing" microelectromechanical components and systems inside a protective housing. Combining engineering and manufacturing technologies, it converts a micromachined structure or system into a *useful* assembly that can *safely* and *reliably* interact with its surroundings. The definition is broad because each application is unique in its packaging requirements. In the integrated circuit industry, electronic packaging must provide reliable dense interconnections to the multitude of high-frequency electrical signals, as well as extract excessive heat from the chips. By contrast, MEMS packaging must account for a far more complex and diverse set of parameters. It must first protect the micromachined parts in broad-ranging environments; it must also provide interconnects to electrical signals and, in most cases, access to and interaction with the external environment. For example, the packaging of a pressure sensor must ensure that the sensing device is in intimate contact with the pressurized medium yet protected from exposure to any harmful substances in this medium. Moreover, packaging of valves must provide both electrical and fluid interconnects, and packaging of lasers must allow for optical fibers. As a consequence of these diverse requirements, standards for MEMS packaging lack, and designs often remain proprietary to companies. Invariably, the difficulty and failure in adopting standards implies that packaging will remain engineering-resource intensive and thus will continue to carry rather high fixed costs.

Packaging is a necessary "evil." Its relatively large dimensions tend to dilute the small-size advantage of MEMS. It is also expensive: the cost of packaging tends to be significantly larger than the cost of the actual micromachined components. It is not unusual that the packaging content is responsible for 75% to 95% of the overall cost of a microelectromechanical component or system. These factors, prevalent in the early days of electronic integrated circuits, contributed towards large-scale integration in that industry in order to minimize the impact of packaging on overall cost, size, and performance. High-density packaging methods, such as surface mount technologies (SMT), are today at the core of advancements in electronic packaging. By contrast, the evolution of MEMS packaging is slow and centers largely on borrowing from the integrated circuit and other industries in an effort to benefit from the existing vast body of knowledge. Whether sophisticated packaging

technologies will penetrate MEMS remains to be seen, but if they do they will certainly have to rely on serious market incentives, in particular high-volume applications, and on a minimum level of technology standardization.

The field of packaging is so broad in scope that one can only hope to present here a brief introduction of the fundamentals (see Figure 8.1), especially as they relate to the various structures and systems introduced in the previous chapters. Such an accomplishment is made more difficult by the proprietary nature of most package designs.

Key Design and Packaging Considerations

Designing packages for micromachined sensors and actuators involves taking into account a number of important factors. Some are shared with the packaging of electronic integrated circuits, but many are specific to the application. These factors also bear significance on the design of the micromachined components themselves. As a result, the design of the package and of the micromachined structures must commence and evolve together; it would be naïve to believe they can be separated. The following are critical factors and considerations frequently encountered in MEMS packaging.

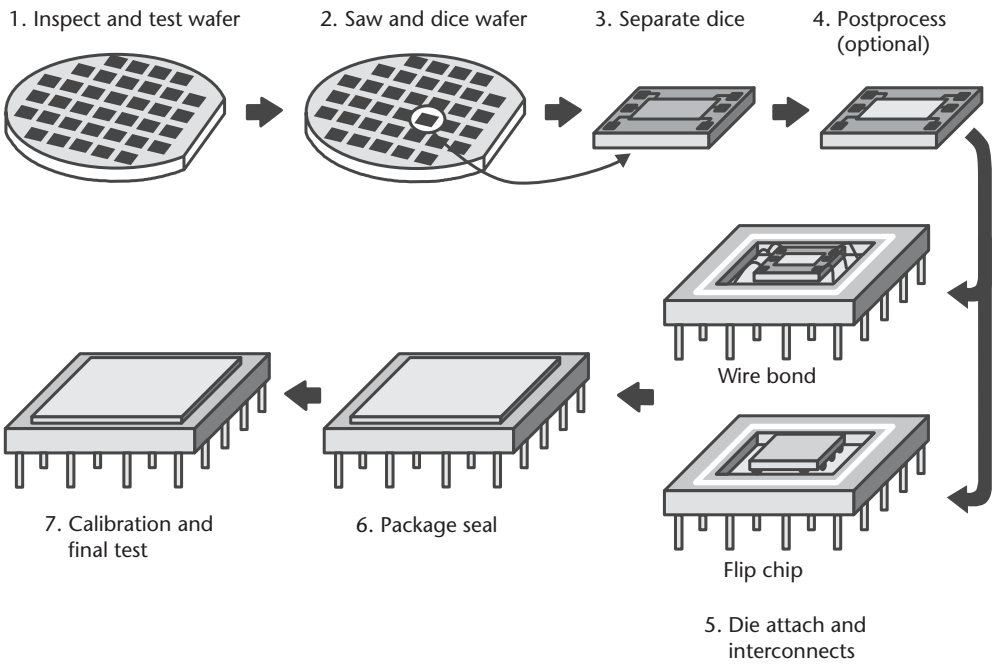


Figure 8.1 Illustration of a simplified process flow for MEMS packaging. Upon completion of wafer-level fabrication, inspection and first tests take place. The wafer is then mounted on a special sticky tape and sawed. The individual dice are separated. Some post processing, such as removal of a sacrificial layer, may occur at this point. One die or many dice are attached to a ceramic, a metal header, or a premolded plastic lead frame. Electrical interconnects are made by wire bonding, flip chip, or another method. A ceramic, metal, glass, or plastic cap seals the assembly. Alternatively, the die or dice are attached to a metal lead frame. After the electrical interconnects are made, plastic is molded over the assembly. A final test and calibration conclude the process. This simple process does not allow for fluidic or optical connections.

Wafer or Wafer-Stack Thickness

Standards in the electronic integrated-circuit industry dictate specific thicknesses for silicon wafers depending on their diameters. For example, a standard 100-mm (4-in) diameter silicon wafer polished on one side has a nominal thickness of 525 μm. The standard thickness increases to 650 μm for 150-mm (6-in) diameter wafers. Wafers polished on both sides are normally thinner. Glass substrates are at least 250 μm (10 mils) thick. Often, a stack of bonded silicon or glass wafers can have a total thickness exceeding 1 mm, posing significant challenges for packaging facilities. In some cases, it becomes outright impossible to accommodate such large thicknesses. Proper communication of the thickness to the parties responsible for packaging is imperative in order to minimize disruptions to the assembly line and avoid unnecessary delays.

Wafer Dicing Concerns

A key highlight of MEMS technology is the batch fabrication aspect—hundreds and thousands of identical structures or microsystems are fabricated simultaneously on the same wafer. Dicing separates these structures into individual components (dice) that can be later packaged. A diamond or carbide saw blade, approximately 50 to 250 μm wide, spins at high speed and cuts through the substrate that is normally mounted and held in position on a colored "sticky tape" known as dicing tape. Water flows continuously during sawing to cool the blade. Dicing is a harsh process conducted in an unclean environment and subjects the microstructures to strong vibrations and flying debris. Retaining the integrity and cleanliness of the microstructures requires protecting the sensitive components from particulates and liquids as well as ensuring that they can survive all of the shaking.

Each MEMS design merits its own distinctive approach on how to minimize the adverse effects of dicing. In surface-micromachined MEMS, such as the accelerometer from Analog Devices, protection can mean, for example, forming shallow dimples in the dicing tape and mounting the wafer upside down such that the sensitive micromechanical structures face toward and are aligned with the dimples. Alternatively, it is possible to perform the final sacrificial etch (see Chapter 3) *after* the dicing is complete. While this *postprocess* approach ensures that there are no free mechanical structures during the dicing, it implies that the microstructures must be freed on each individual die, thus sacrificing batch fabrication for mechanical integrity. This naturally increases the final fabrication cost. The fabrication process of the Texas Instruments, Digital Mirror Device (DMD) follows this approach. The DMD arrays are diced first, then the organic sacrificial layer on each individual die is subsequently etched in oxygen plasma. Because the rumored selling price for each DMD is in the hundreds of dollars, this method may be economically justified, but accelerometers intended for the automotive market command prices of a few dollars at most with little margin to allocate to the dicing process.

The reader will observe in Chapters 4 through 7 a number of designs incorporating bonded caps or covers made of silicon and occasionally glass, whose sole purpose is to protect the sensitive micromechanical structures. These become, after the completion of the cap, fully embedded inside an all-micromachined housing—a first-level package. For example, the yaw-rate sensor from Robert Bosch GmbH includes a silicon cover that protects the embedded microstructures during dicing,

provided the vibrations are not sufficiently large to cause damage. In addition to mechanical protection, an electrically grounded cover also shields against electromagnetic interference (EMI). Naturally, the cap approach is not suitable for sensors, such as pressure or flow sensors, or actuators that require direct and immediate contact with their surrounding environments.

Thermal Management

The demands on thermal management can be very diverse and occasionally conflicting depending on the nature of the application. The main role of thermal management for electronic packaging is to cool the integrated circuit during operation [1]. A modern microprocessor containing millions of transistors and operating at a few gigahertz can consume tens of watts. By contrast, the role of thermal management in MEMS includes the cooling of heat-dissipating devices and, especially, thermal actuators, but it also involves understanding and accounting for the sources of temperature fluctuations that may adversely affect the performance of a sensor or actuator. As such, thermal management is performed at two levels: the die level and the package level.

Thermal analysis is analogous to understanding electrical networks. This is not surprising because of the dual nature of heat and electricity—voltage, current, and electrical resistance are dual to temperature, heat flux, and thermal resistance, respectively. A network of resistors is an adequate first-order model to understand heat flow and nodal temperatures. The thermal resistance, θ, of an element is equal to the ratio of the temperature difference across the element to the heat flux—this is equivalent to Ohm's law for heat flow. For a simple slab of area A and length l, θ equals $l/(\kappa A)$ where κ is the thermal conductivity of the material (see Figure 8.2).

The nature of the application severely influences the thermal management at the die level. For example, in typical pressure sensors that dissipate a few milliwatts over

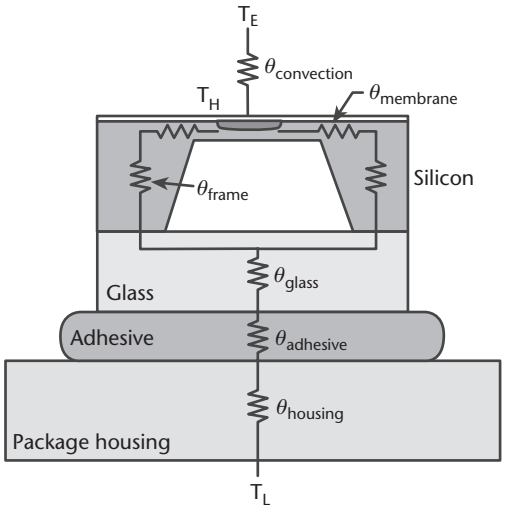


Figure 8.2 Components of thermal resistance for a hypothetical microstructure, including a heat-producing element at temperature T_H, embedded in a suspended membrane. The device is assembled within a housing maintained at a low temperature, T_L. The temperature of the surrounding environment is T_E.

an area of several square millimeters, the role of thermal management is to ensure long-term thermal stability of the piezoresistive sense elements by verifying that no thermal gradients arise within the membrane. The situation becomes more complicated if any heat-dissipating elements are positioned on very thin membranes, increasing the effective thermal resistance to the substrate and the corresponding likelihood of temperature fluctuations. Under some circumstances, maintaining an element at a constant temperature above ambient brings performance benefits. One example is the mass-flow sensor from Honeywell (see Chapter 4).

Thermal management at the package level must take into account all of the thermal considerations of the die level. In the case of the mass-flow sensor, it is imperative that the packaging does not interfere with the die-level thermal isolation scheme. In the example of the infrared imager also from Honeywell (see Chapter 5), the package housing needs to hold a permanent vacuum to eliminate convective heat loss from the suspended sensing pixels.

Thermal actuators can dissipate significant power. It can take a few watts for a thermal actuator to deliver a force of 100 mN with a displacement of 100 μm. With efficiencies typically below 0.1%, most of the power is dissipated as heat that must be removed through the substrate and package housing. In this case, thermal management shares many similarities with the thermal management of electronic integrated circuits. This is a topic that is thoroughly studied and discussed in the literature [1].

Metals and some ceramics make excellent candidate materials for the package housing because of their high thermal conductivity. To ensure unimpeded heat flow from the die to the housing, it is necessary to select a die-attach material that does not exhibit a low thermal conductivity. This may exclude silicones and epoxies and instead favor solder-attach methods or silver-filled epoxies, polyimides, or glasses. A subsequent section in this chapter explores various die-attach techniques. Naturally, a comprehensive thermal analysis should take into account all mechanisms of heat loss, including loss to fluid in direct contact with the actuator.

Stress Isolation

The previous chapters described the usefulness of piezoresistivity and piezoelectricity to micromachined sensors. By definition, such devices rely on converting mechanical stress to electrical energy. It is then imperative that the piezoresistive or piezoelectric elements are not subject to mechanical stress of undesirable origin and extrinsic to the parameter that needs to be sensed. For example, a piezoresistive pressure sensor gives an incorrect pressure measurement if the package housing subjects the silicon die to stresses. These stresses need only be minute to have a catastrophic effect because the piezoresistive elements are extremely sensitive to stress. Consequently, sensor manufacturers take extreme precautions in the design and implementation of packaging. The manufacture of silicon pressure sensors, especially those designed to sense low pressures ($<$100 kPa), includes the anodic bonding of a thick ($>$1 mm) Pyrex glass substrate with a coefficient of thermal expansion matched to that of silicon. The glass improves the sensor's mechanical rigidity and ensures that any stresses between the sensor and the package housing are isolated from the silicon piezoresistors.

Another serious effect of packaging on stress-sensitive sensors is long-term drift resulting from slow creep in the adhesive or epoxy that attaches the silicon die to the package housing. Modeling of such effects is extremely difficult, leaving engineers with the task of constant experimentation to find appropriate solutions. This illustrates the type of "black art" that exists in the packaging of sensors and actuators, and it's a reason companies do not disclose their packaging secrets.

Protective Coatings and Media Isolation

Sensors and actuators coming into intimate contact with external media must be protected against adverse environmental effects, especially if the devices are subject to long-term reliability concerns. This is often the case in pressure or flow sensing, where the medium in contact is other than dry air. For example, sensors for automotive applications must be able to withstand salt water and acid rain pollutants (e.g., SO_x, NO_x). In home appliances (white goods), sensors may be exposed to alkali environments due to added detergents in water. Even humidity can cause severe corrosion of sensor metallization, especially aluminum.

In many instances of mildly aggressive environments, a thin conformal coating layer is sufficient protection. A common material for coating pressure sensors is parylene (poly(*p*-xylylene) polymers) [2, 3] (see Table 8.1). It is normally deposited using a near-room-temperature chemical vapor deposition process. The deposited film is conformal covering the sensor element and exposed electrical wires. It is resistant to automotive exhaust gases, fuel, salt spray, water, alcohol, and many organic solvents. However, extended exposure to highly acidic or alkali solutions ultimately results in the failure of the coating.

Recent studies suggest that silicon carbide may prove to be an adequate coating material to protect MEMS in very harsh environments [4]. Silicon carbide deposited in a plasma-enhanced chemical vapor deposition (PECVD) system by the pyrolysis of silane (SiH_4) and methane (CH_4) at 300°C proved to be an effective barrier for protecting a silicon pressure sensor in a hot potassium hydroxide solution, which is a highly corrosive chemical and a known etchant of silicon. However, much

Table 8.1 Material Properties for Three Types of Parylene Coatings*

Property	*Parylene-N*	*Parylene-C*	*Parylene-D*
Density (g/cm^{-3})	1.110	1.289	1.418
Tensile modulus (GPa)	2.4	3.2	2.8
Permittivity	2.65	3.15	2.84
Volume resistivity (Ω•cm) at 23°C, 50% RH	1.2×10^{17}	8.8×10^{16}	1.2×10^{17}
Refractive index	1.661	1.639	1.669
Melting point (°C)	410	290	380
Coefficient of expansion (10^{-6} /K)	69	35	<80
Thermal conductivity (W/m•K)	0.12	0.082	—
Maximum water absorption (%)	0.01	0.06	<0.1
Gas permeability (amol/Pa•s•m)			
N_2	15.4	2.1	9.0
CO_2	429.0	15.4	26.0
SO_2	3,790.0	22.0	9.53

*They are stable at cryogenic temperatures to over 125°C [2].

development remains to be done to fully characterize the properties of silicon carbide as a coating material.

For extreme environments such as in applications involving heavy industries, aerospace, or oil drilling, special packaging is necessary to provide adequate protection to the silicon microstructures. If the silicon parts need not be in direct contact with the surrounding environment, then a metal or ceramic hermetic package may be sufficient. This is adequate for accelerometers, for example, but inappropriate for pressure or flow sensors. Such devices must be isolated from direct exposure to their surrounding media and yet continue to measure pressure or flow rate. Clever media-isolation schemes for pressure sensors involve immersing the silicon microstructure in special silicone oil with the entire assembly contained within a heavy-duty stainless-steel package. A flexible stainless-steel membrane allows the transmission of pressure through the oil to the sensor's membrane. Media-isolated pressure sensors are discussed in further detail later in this chapter.

Media-isolation can be more difficult to achieve in certain applications. For instance, there are numerous demonstrations of optical microspectrometers capable of detecting SO_x and NO_x, two components of smog pollution. But incorporating these sensors into the tail pipe of an automobile has proven to be of great difficulty because the sensor must be isolated from the harsh surrounding environment, yet light must reach the sensor. A transparent glass window is not adequate because of the long-term accumulation of soot and other carbon deposits.

Hermetic Packaging

A hermetic package is theoretically defined as one that prevents the diffusion of helium. For small-volume packages ($<0.40\ cm^{-3}$), the leak rate of helium must be lower than 5×10^{-8} atm•cm^3/s. In practice, it is always understood that a hermetic package prevents the diffusion of moisture and water vapor through its walls. A hermetic package must be made of metal, ceramic, or millimeter-thick glass. Silicon also qualifies as a hermetic material. Plastic and organic-compound packages, on the other hand, may pass the strict helium leak rate test, but they allow moisture into the package interior over time; hence, they are not considered hermetic. Electrical interconnections through the package must also conform to hermetic sealing. In ceramic packages, metal pins are embedded and brazed within the ceramic laminates. For metal packages, glass firing yields a hermetic glass-metal seal.

A hermetic package significantly increases the long-term reliability of electrical and electronic components. By shielding against moisture and other contaminants, many common failure mechanisms including corrosion are simply eliminated. For example, even deionized water can leach out phosphorous from low-temperature oxide (LTO) passivation layers to form phosphoric acid that, in turn, etches and corrodes aluminum wiring and bond pads. The interior of a hermetic package is typically evacuated or filled with an inert gas such as nitrogen, argon, or helium. The DMD from Texas Instruments and the infrared imager from Honeywell, both discussed in a previous chapter, utilize vacuum hermetic packages with transparent optical windows. The package for the DMD even includes a getter to absorb any residual moisture.

Calibration and Compensation

The performance characteristics of precision sensors, especially pressure, flow, acceleration, and yaw-rate sensors, often must be calibrated in order to meet the required specifications. Errors frequently arise due to small deviations in the manufacturing process. For example, the sensitivity of a pressure sensor varies with the square of the membrane thickness. A typical error of ±0.25 μm on a 10-μm thick membrane produces a ±5% error in sensitivity that must be often trimmed to less than ±1%. Additionally, any temperature dependence of the output signal must be compensated.

One compensation and calibration scheme utilizes a network of laser-trimmed resistors with near-zero TCR to offset errors in the sensor [5]. The approach employs all-passive components and is an attractive low-cost solution. The resistors can be either thin film (<1 μm thick) or thick film (~ 25 μm thick) [6] and are trimmed by laser ablation. Thin-film resistors, frequently used in analog integrated circuits such as precision operational amplifiers, are sputtered or evaporated directly on the silicon die and are usually made of nickel-chromium or tantalum-nitride. These materials have a sheet resistance of about 100 to 200Ω per square, and a very low TCR of ±0.005% per degree Celsius. Nickel-chromium can corrode if not passivated with quartz or silicon monoxide (SiO), but tantalum nitride self passivates by baking in air for a few minutes. Thick-film resistors, by contrast, are typically fired on thick ceramic substrates and consist of chains of metal-oxide particles embedded in a glass matrix. Ruthenium dioxide (RuO_2) and bismuth ruthenate ($BiRu_2O_7$) are examples of active metal oxides. Blending the metal oxides with the glass in different proportions produces sheet resistances with a range of values from 10 to 10^6 Ω per square. Their TCR is typically in the range of ±0.01% per degree Celsius. Trimming using a neodymium-doped yttrium-aluminum-garnet (Nd:YAG) laser at a wavelength of 1.06 μm produces precise geometrical cuts in the thin- or thick-film resistor, hence adjusting its resistance value. The laser is part of a closed-loop system that continuously monitors the value of the resistance and compares it to a desired target value.

Laser ablation is also useful to calibrate critical mechanical dimensions by direct removal of material. For instance, a laser selectively ablates minute amounts of silicon to calibrate the two resonant modes of the Daimler Benz tuning fork yaw-rate sensor (see Chapter 4). Laser ablation can also be a useful process to precisely calibrate the flow of a liquid through a micromachined channel. For some drug delivery applications, such as insulin injection, the flow must be calibrated to within ±0.5%. Given the inverse cubic dependence of flow resistance on channel depth, this translates to an etch depth precision of better than ±0.17%, equivalent to 166 nm in a 100-μm deep channel. This is impossible to achieve using most, if not all, silicon-etching methods. A laser ablation step can control the size of a critical orifice under closed-loop measurement of the flow to yield the required precision.

As the integration of circuits and sensors becomes more prevalent, the trend has been to perform, when possible, calibration and compensation electronically. Many modern commercial sensors, including pressure, flow, acceleration, and yaw-rate sensors, now incorporate application-specific integrated circuits (ASICs) to calibrate the sensor's output and compensate for any errors. Correction coefficients are stored in on-chip permanent memory such as EEPROM.

The need to calibrate and compensate extends beyond conventional sensors. For example, the infrared imaging array from Honeywell must calibrate each individual pixel in the array and compensate for any manufacturing variations across the die. The circuits perform this function using a shutter: The blank scene, that is the collected image while the shutter is closed, incorporates the variation in sensitivity across the array; while the shutter is open, the electronic circuits subtract the blank-scene image from the active image to yield a calibrated and compensated picture.

Die-Attach Processes

Subsequent to dicing of the substrate, each individual die is mounted inside a package and attached (bonded) onto a platform made of metal or ceramic, though plastic is also possible under limited circumstances. Careful consideration must be given to die attaching because it strongly influences thermal management and stress isolation. Naturally, the bond must not crack over time nor suffer from creep—its reliability must be established over very long periods of time. The following section describes die-attach processes common in the packaging of silicon micromachined sensors and actuators. These processes were largely borrowed from the electronics industry.

Generally, die-attach processes employ either metal alloys or organic or inorganic adhesives as intermediate bonding layers [7, 8]. Metal alloys comprise of all forms of solders, including eutectic and noneutectic (see Table 8.2). Organic adhesives consist of epoxies, silicones, and polyimides. Solders, silicones, and epoxies are vastly common in MEMS packaging. Inorganic adhesives are glass matrices

Table 8.2 Properties of Some Eutectic and Noneutectic Solders

	Alloy	*Liquidus (°C)*	*Solidus (°C)*	*Ultimate Tensile Strength (MPa)*	*Uniform Elongation (%)*	*Creep Resistance*
Noneutectic	60%In 40%Pb	185	174	29.58	10.7	Moderate
	60%In 40%Sn	122	113	7.59	5.5	Low—soft alloy
	80%In 15%Pb 5%Ag	154	149	17.57	—	Low
	80%Sn 20%Pb	199	183	43.24	0.82	Moderate
	25%Sn 75%Pb	266	183	23.10	8.4	Poor
	5%Sn 95%Pb	312	308	23.24	26	Moderate to high
	95%Sn 5%Sb	240	235	56.20	1.06	High
Eutectic	97%In 3%Ag	143	143	5.50	—	Low—soft alloy
	96.5%Sn 3.5%Ag	221	221	57.65	0.69	High
	42%Sn 58%Bi	138	138	66.96	1.3	Moderate —brittle alloy
	63%Sn 37%Pb	183	183	35.38	1.38	Moderate
	1%Sn 97.5%Pb 1.5%Ag	309	309	38.48	1.15	Moderate
	88%Au 12%Ge	356	356	—	—	Moderate
	96.4%Au 3.6%Si	370	370	—	—	Moderate

(*Source:* [7].)

embedded with silver and resin and are mostly used in the brazing of pressed ceramic packages (e.g., CERDIP type and CERQUAD type) in the integrated circuits industry. Their utility for die-attach may be limited because of the high-temperature (400°C) glass seal and cure operation.

The choice of a solder alloy depends on it having a suitable melting temperature as well as appropriate mechanical properties. A solder firmly attaches the die to the package and normally provides little or no stress isolation when compared to organic adhesives. The large mismatch in the coefficients of thermal expansion with silicon or glass results in undesirable stresses that can cause cracks in the bond. However, the bond is very robust and can sustain large normal pull forces on the order of 5,000 N/cm^2.

Most common solders are binary or ternary alloys of lead (Pb), tin (Sn), indium (In), antimony (Sb), bismuth (Bi), or silver (Ag) (see Figure 8.3). Solders can be either hard or soft. Hard solders (or brazes) melt at temperatures near or above 500°C and are used for lead and pin attachment in ceramic packages. By contrast, soft solders melt at lower temperatures, and, depending on their composition, they are classified as eutectic or noneutectic. Eutectic alloys go directly from liquid to solid phase without an intermediate paste-like state mixing liquid and solid—effectively, eutectic alloys have identical solidus and liquidus temperatures. They have the lowest melting points of alloys sharing the same constituents and tend to be more rigid with excellent shear strength.

Silicon and glass cannot be directly soldered to and thus must be coated with a thin metal film to wet the surface. Platinum, palladium, and gold are good choices, though gold is not as desirable with tin-based solders because of leaching. Leaching is the phenomenon by which metal is absorbed into the solder to an excessive degree causing intermetallic compounds detrimental to long-term reliability—gold or silver will dissolve into a tin-lead solder within a few seconds. Typically, a thin (<50 nm) layer of titanium is first deposited on the silicon to improve adhesion, followed by the deposition of a palladium, platinum, or nickel layer, a few hundred nanometers thick—this layer also serves as a diffusion barrier. A subsequent flash

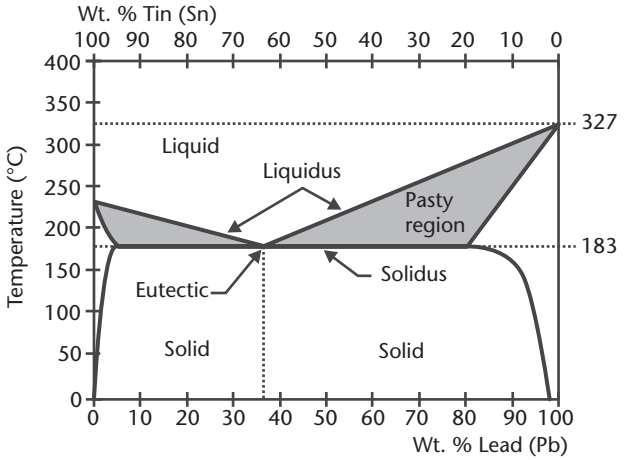


Figure 8.3 Phase diagram of lead-tin solder alloys. The eutectic point corresponds to a lead composition of 37% by weight [7].

deposition of very thin gold improves surface wetting. Immersing the part in flux (an organic acid) removes metal oxides and furnishes clean surfaces. In a manufacturing environment, the solder paste is either dispensed through a nozzle or screen printed on the package substrate, and the die is positioned over the solder. Heating in an oven or by direct infrared radiation melts the solder, dissolving in the process a small portion of the exposed thin metal surfaces. When the solder cools, it forms a joint bonding the die to the package. Melting in nitrogen or in forming gas prevents oxidation of the solder.

Organic adhesives are attractive alternatives to solder because they are inexpensive, easy to automate, and they cure at lower temperatures. The most widely used are epoxies and silicones, including room-temperature vulcanizing (RTV) rubbers. Epoxies are thermosetting (i.e., cross linking when heated) plastics with cure temperatures varying between room temperature and 175°C. Filled with silver or gold, they become thermally and electrically conductive, but not as conductive as solder. Electrically nonconductive epoxies may incorporate particles of aluminum oxides, beryllium oxides, or magnesium oxides for improved thermal conductivity. RTV silicones come in a variety of specifications for a wide range of applications from construction to electronics. For example, the Dow Corning® 732 is a multipurpose silicone that adheres well to glass, silicon, and metal, with a temperature rating of –65°C to 232°C [9]. Most RTV silicones are one part condensation-curing compounds, curing at room temperature in air while outgassing a volatile reaction product, such as acetic acid. Another class of RTVs, however, is addition-cure RTVs, which do not outgas, making them suitable for many optical applications. Unlike epoxies, they are soft and are excellent choices for stress relief between the package and the die. The operating temperature for most organic adhesives is limited to less than 200°C; otherwise, they suffer from structural breakdown and outgassing.

Epoxies and RTV silicones are suitable for automated manufacturing. As viscous pastes, they are dispensed by means of nozzles at high rates or screen printed. The placement of the die over the adhesive may also be automated by using *pick-and-place* robotic stations employing pattern recognition algorithms for accurate positioning of the die.

Wiring and Interconnects

With the advent of microfluidic components and systems, the concept of interconnects is now more global, simultaneously incorporating electrical and fluid connectivity. Electrical connectivity addresses the task of providing electrical wiring between the die and electrical components external to it. The objective of fluid connectivity is to ensure the reliable transport of liquids and gases between the die and external fluid control units.

Electrical Interconnects

Wire Bonding

Wire bonding is unquestionably the most popular technique to electrically connect the die to the package. The free ends of a gold or aluminum wire form low-resistance

(ohmic) contacts to aluminum bond pads on the die and to the package leads (terminals). Bonding gold wires tends to be easier than bonding aluminum wires.

Thermosonic gold bonding is a well-established technique in the integrated circuit industry, simultaneously combining the application of heat, pressure, and ultrasonic energy to the bond area. Ultrasound causes the wire to vibrate, producing localized frictional heating to aid in the bonding process. Typically, the gold wire forms a *ball bond* to the aluminum bond pad on the die and a *stitch bond* to the package lead. The "ball bond" designation follows after the spherical shape of the wire end as it bonds to the aluminum. The stitch bond, in contrast, is a wedge-like connection as the wire is pressed into contact with the package lead (typically gold or silver plated). The temperature of the substrate is usually near 150°C, below the threshold of the production of gold-aluminum intermetallic compounds that cause bonds to be brittle. One of these compounds (Au_5Al_{12}) is known as *purple plague* and is responsible for the formation of voids—the Kirkendall voids—by the diffusion of aluminum into gold. Thermosonic gold bonding can be automated using equipment commercially available from companies such as Kulicke and Soffa Industries, Inc., of Willow Grove, Pennsylvania.

Bonding aluminum wires to aluminum bond pads is also achieved with ultrasonic energy but without heating the substrate. In this case, a stitch bond works better than a ball bond, but the process tends to be slow. This makes bonding aluminum wires economically not as attractive as bonding gold wires. However, gold wires are difficult to obtain with diameters above 50 μm (2 mils), which makes aluminum wires, available in diameters up to 560 μm (22 mils), the only solution for high-current applications (see Table 8.3).

The thermosonic ball bond process begins with an electric discharge or spark to melt the gold and produce a ball at the exposed wire end (see Figure 8.4). The tip—or capillary—of the wire-bonding tool descends onto the aluminum bond pad, pressing the gold ball into bonding with the bond pad. Ultrasonic energy is simultaneously applied. The capillary then rises and the wire is fed out of it to form a loop as the tip is positioned over the package lead—the next bonding target. The capillary is lowered again, deforming the wire against the package lead into the shape of a wedge—the stitch bond. As the capillary rises, special clamps close onto the wire, causing it to break immediately above the stitch bond. The size of the ball dictates a minimum in-line spacing of approximately 100 μm between adjacent bond pads on the die. This spacing decreases to 75 μm for stitch bonding.

Table 8.3 Recommended Maximum Current in Gold and Aluminum Bond Wires

		Maximum current (A)	
Material	*Diameter (μm)*	*Length <1 mm*	*Length < 1 mm*
Gold	25	0.95	0.65
	50	2.7	1.8
Aluminum	25	0.7	0.5
	50	2	1.4
	125	7.8	5.4
	200	15.7	10.9
	300	28.9	20
	380	40.4	27.9
	560	71.9	49.6

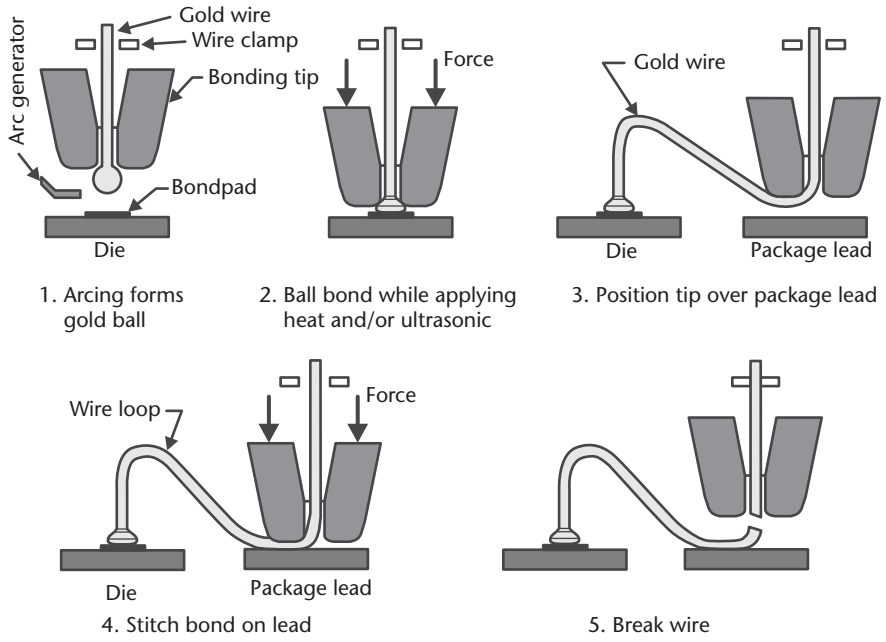


Figure 8.4 Illustration of the sequential steps in thermosonic ball and stitch bonding. The temperature of the die is typically near 150°C. Only the tip of the wire-bonding tool is shown [10].

The use of wire bonding occasionally runs into serious limitations in MEMS packaging. For instance, the applied ultrasonic energy, normally at a frequency between 50 and 100 kHz, may stimulate the oscillation of suspended mechanical microstructures. Unfortunately, many micromachined structures coincidentally have resonant frequencies in the same range, increasing the risk of structural failure during wire bonding.

Flip Chip

Flip-chip bonding [11], as its name implies, involves bonding the die, top face down, on a package substrate (see Figure 8.5). Electrical contacts are made by means of plated solder bumps between bond pads on the die and metal pads on the package substrate. The attachment is intimate with a relatively small spacing (50 to 200 μm) between the die and the package substrate. Unlike wire bonding which requires the bond pads to be positioned on the periphery of the die to avoid crossing wires, flip chip allows the placement of bond pads over the entire die (area arrays), resulting in a significant increase in density of input/output (I/O) connections—up to 700 simultaneous I/Os. Additionally, the effective inductance of each interconnect is miniscule because of the short height of the solder bump. The inductance of a single solder bump is less than 0.05 nH, compared to 1 nH for a 125-μm-long and 25-μm-diameter wire. It becomes clear why the integrated circuit industry has adopted flip chip for high-density, fast electronic circuits.

What makes flip-chip bonding attractive to the MEMS industry is its ability to closely package a number of distinct dice on one single package substrate with multiple levels of embedded electrical traces. For instance, one can use flip-chip bonding

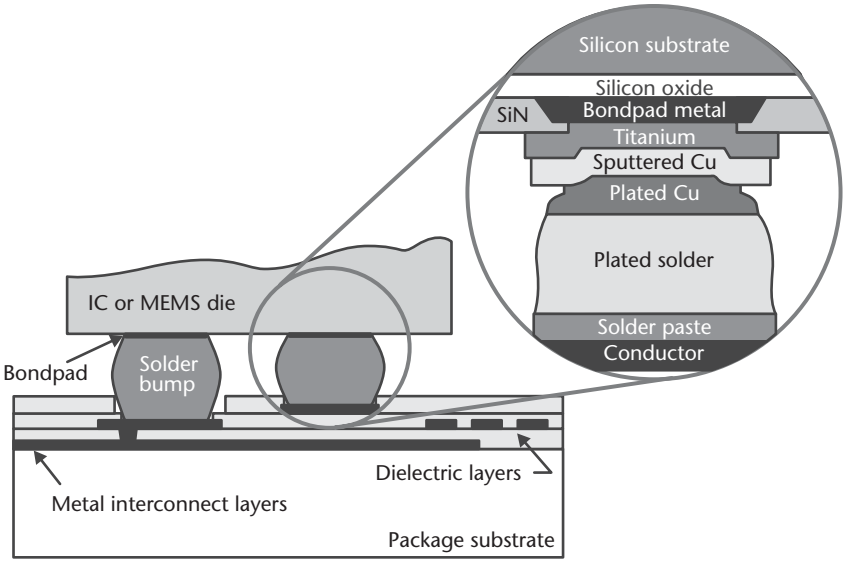


Figure 8.5 Flip-chip bonding with solder bumps.

to electrically connect and package three accelerometer dice, a yaw-rate sensing die, and an electronic ASIC onto one ceramic substrate to build a fully self-contained navigation system. This type of hybrid packaging produces complex systems, though each individual component in itself may not be as complex. Clearly, a similar system can be built with wire bonding, but its area usage will not be as efficient and its reliability may be questionable, given the large number of gold wires within the package (note that each suspended gold wire is in essence an accelerometer, subject to deflections and potential shorting).

Additional fabrication steps are required to form the solder bumps over the die. A typical process involves the sputtering of a titanium layer over the bond pad metal (e.g., aluminum) to promote adhesion, followed by the sputtering of copper. Patterning and etching of the titanium and copper defines a pedestal for the solder bump. A thicker layer of copper is then electroplated. Finally, the solder bump, typically a tin-lead alloy, is electroplated over the copper. Meanwhile, in a separate preparation process, solder paste is screen printed on the package substrate in patterns corresponding to the landing sites of the solder bumps. Automated pick-and-place machines position the die, top face down, and align the bond pads to the solder-paste pattern on the package substrate. Subsequent heating in an oven or under infrared radiation melts the solder into a columnar, smooth, and shiny bump. Surface tension of the molten solder is sufficient to correct for any slight misalignment during the die-positioning process. If desired, a final underfill step fills the void space between the die and the package substrate with epoxy. An optional silicone or parylene conformal coat protects the entire assembly.

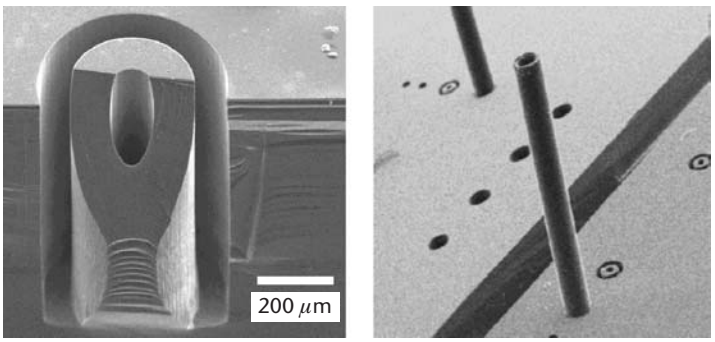
Flip chip may not be compatible with the packaging of MEMS with microstructures exposed to the open environment. For instance, there is a risk of damaging the thin diaphragm of a pressure sensor during a flip-chip process. By contrast, a capped device such as the Bosch yaw-rate sensor (see Chapter 4) can take full advantage of flip-chip technology.

Microfluidic Interconnects

All advances in electrical interconnect technology derive from the packaging requirements of the integrated circuit industry, but that is not the case for fluidic interconnects. These are required to package microfluidic devices such as micropumps and microvalves. No standards exist simply because the field remains in its infancy and few microfluidic devices are commercially available. Sadly, most microfluidic interconnect schemes remain at the level of manually inserting a capillary into a silicon cavity or via-hole and sealing the assembly with silicone or epoxy (see, for example, the PCR thermal cycler in Chapter 6). These are suitable methods for laboratory experimentation but will not meet the requirements of automated manufacturing (see Figure 8.6).

Future fluid packaging schemes amenable to high-volume manufacturing would have to rely on simplified fluid interconnects. For example, fluid ports in a silicon die could be aligned directly to ports in a ceramic or metal manifold. The silicon die can be attached by any of the die-attach methods described earlier. Under such a scheme, it becomes possible to envisage systems with fluid connectivity on one side of the die and electrical connectivity on the opposite side. This would enhance long-term reliability by separating fluid flow from electrical wiring.

Researchers at Abbott Laboratories of Abbott Park, Illinois, demonstrated a hybrid packaging approach incorporating a complex manifold in acrylic (e.g., Plexiglas™) [13]. These are large boards, many centimeters in size, with multiple levels of channels and access vias, all made in plastic. The channels are formed by laminating and bonding layers of thermoplastics into which trenches had been preformed. The plastic board becomes equivalent to a *fluid printed-circuit board* onto which surface fluid components are attached and wired. These components need not necessarily be micromachined. For example, the board could hold a silicon pressure or flow sensor in proximity of a miniature solenoid valve. Much of the technology for fluid interconnects remains under development. New markets and applications will undoubtedly drive engineers to contrive innovative but economically justifiable solutions.



(a) (b)

Figure 8.6 (a) A photograph of a fluid interconnect etched in silicon using DRIE. Fluid flows through a central orifice leading into a channel embedded within the silicon substrate. The precise outer trench provides mechanical support to tightly hold a capillary in position. (b) A photograph of a capillary inserted into an intact fluid port. (*Courtesy of:* GE NovaSensor of Fremont, California [12].)

Optical Interconnects

Optical interconnecting is generally understood as the active field of research that aims to develop very fast chip-to-chip transmission rates for high-speed computation. For the packaging of MEMS components in photonic applications, optical interconnects are simpler in nature, generally entailing coupling light in and out of an optical fiber without a noticeable loss (typically <0.1 dB). The laser and optoelectronic industry makes extensive use of fiber interconnects for the packaging and manufacture of their products. Companies packaging optical MEMS components have largely borrowed these established packaging designs and methods for their applications. One such design is the ubiquitous gold-plated butterfly package [14], which includes electrical pins in a winged construction with an allowance for a fiber connection [see Figure 8.7(a)].

Establishing an optical connection through the walls of the butterfly package involves positioning an optical fiber inside a feed-through tube and aligning it relative to optical elements inside the package. This alignment step is critical and is often completed actively *in situ*, with light propagating through the fiber during the assembly to guarantee maximum optical coupling between the fiber and the components inside the package [15]. A hermetic seal of the feed through is also necessary because the entire butterfly package is hermetically sealed with a top cover at the end of the assembly process. Hermetic sealing of optical components is a pillar of high-reliability packaging required under the Telcordia® standards of the

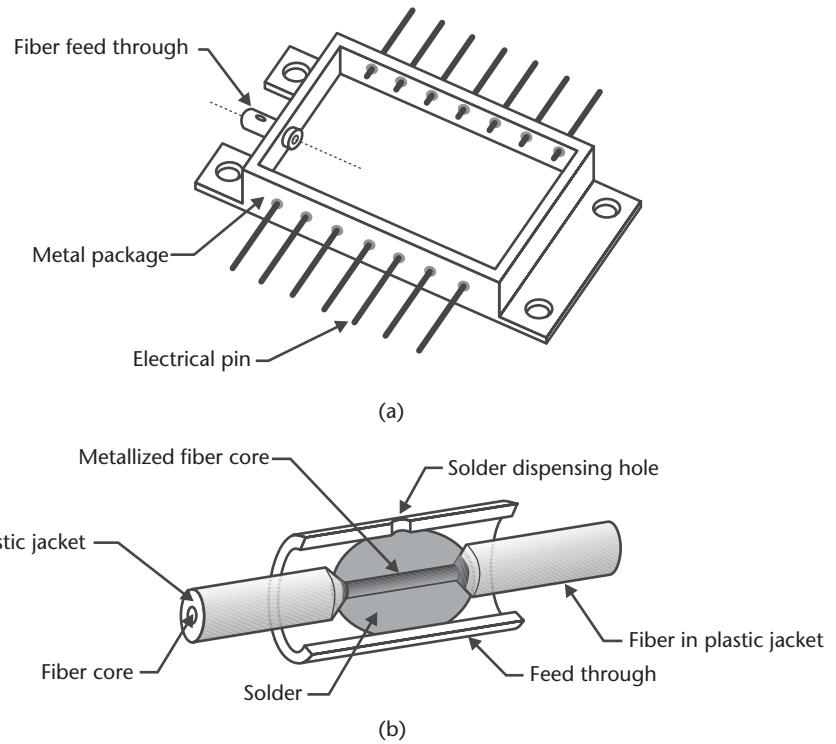


Figure 8.7 (a) A schematic of the gold-plated butterfly package, commonly used in the packaging of fiber-based optical components; and (b) an illustration showing a fiber soldered inside the package feed through. The plastic jacket surrounding the fiber core is stripped and metallized prior to soldering.

telecommunications industry. Solder is a common material to hermetically seal the feed-through tube. The fiber plastic jacket is first stripped over a short distance, exposing the glass core of the fiber, which is then metallized with layers of nickel and gold. The fiber is subsequently inserted into the feed through, and solder is dispensed through a tiny opening in the tube [see Figure 8.7(b)]. Occasionally, an intermediate metal ferrule is used between the fiber and the feed through [16]. Indium-based solders, such as In 97%/Ag 3% or In 80%/Pb 15%/Ag 5%, are common, as they offer good wetting and low melting temperatures ($\leq 150^{\circ}C$) to minimize the risk of damage to the fiber. Because they are soft alloys, they exhibit little stresses at the fiber surface.

Types of Packaging Solutions

In its basic form, a package is a protective housing with an enclosure to hold one or multiple dice forming a complete microelectromechanical device or system. The package provides where necessary electrical, optical, and fluid connectivity between the dice and the external world.

In some cases, it is advantageous to provide a first level of packaging (chip- or die-level encapsulation) to the micromechanical structures and components [17]. This is particularly of interest in applications where the surfaces of the microstructures need not be in direct exposure to liquids or gases. A top silicon cap attached, for example, by silicon fusion bonding can maintain a hermetic seal and hold a vacuum while protecting the sensitive microstructures from damage during saw and assembly. A top cap also allows the use of plastic molding, ubiquitous in low-cost packaging solutions. In this method, molten plastic flows under high pressure, filling the inner cavity of a mold and encapsulating a metal lead frame. The die or capped microstructure rests upon this frame. For example, a crystalline silicon cap protects the sensing elements of the accelerometer from VTI Technologies (see Chapter 4) during molding of the plastic package over the die. Fixed to ground potential, the cap also becomes an effective shield against electromagnetic interference [18].

There are three general categories of widely adopted packaging approaches in MEMS. They are ceramic, metal, and plastic, each with their own merits and limitations (see Table 8.4). For instance, plastic is a low-cost, oftentimes a small-size (surface mount) solution, but it is inadequate for harsh environments. The asking price for a plastic packaged pressure or acceleration sensor is frequently below $5. By contrast, a similar sensor packaged in a hermetic metal housing may cost well over $30. It is not surprising that packaging is what frequently determines economic competitiveness.

Ceramic Packaging

Ceramics are hard and brittle materials made by shaping a nonmetallic mineral, then firing at a high temperature for densification. The vast majority of ceramics are electrical insulators and often are good thermal conductors (see Table 8.5). Ease of shaping along with reliability and attractive material properties (e.g., electrical insulator, hermetic sealing) have made ceramics a mainstay in electronic packaging. They are widely used in multichip modules (MCM) [19] and advanced electronic

Table 8.4 The Diversity of MEMS Packaging Requirements

		Electrical Contacts	*Fluid Ports*	*Media Contact*	*Transparent Window*	*Hermetic Sealing*	*Stress Isolation*	*Heat Sinking*	*Thermal Isolation*	*Calibration and Compensation*	*Types of Packaging*[†]
Sensors	*Pressure*	Yes	Yes	Yes	No	Possibly	Yes	No	No	Yes	P, M, C
	Flow	Yes	Yes	Yes	No	No	No	No	Yes	Yes	P, M, C
	Acceleration	Yes	No	No	No	Yes	Possibly	No	No	Yes	P, M, C
	Yaw rate	Yes	No	No	No	Yes	Possibly	No	No	Yes	P, M, C
	Microphone	Yes	Yes	Yes	No	No	No	No	No	Yes	P, M, C
	Hydrophone	Yes	Yes	Yes	No	Possibly	No	No	No		M, C
Actuators	*Optical switch*	Yes	No	No	Yes	Yes	No	No	No	Yes	M, C
	Display	Yes	No	No	Yes	Yes	No	Possibly	Possibly	No	M, C
	Valve	Yes	Yes	Yes	No	No	No	Possibly	Possibly	Possibly	M, C
	Pump	Yes	Yes	Yes	No	No	No	Possibly	No	Possibly	M, C
	PCR thermal cycler	Yes	Yes	Yes	Possibly	No	No	Possibly	Yes	No	M, C
	Electrophoresis	Yes	Yes	Yes	Yes	No	No	No	No	No	M, C
Passive	*Nozzles*	No	Yes	Yes	No	No	No	No	No	No	P, M, C
	Fluid mixer	No	Yes	Yes	Possibly	No	No	No	No	No	P, M, C
	Fluid amplifier	No	Yes	Yes	No	No	No	No	No	Possibly	M, C

[*] Fluid includes liquid or gas.

[†] P: plastic, M: metal, C: ceramic

packages such as ball grid arrays (BGA) [20]. These same characteristics have extended the utility of ceramics to the packaging of MEMS—many commercially available micromachined sensors use some form of ceramic packaging. Ceramics are completely customizable and allow the formation of through ports and manifolds for the packaging of fluid-based MEMS. Ceramics usually suffer from shrinkage (~13% in the horizontal direction and ~15% in the vertical direction) during firing, which manufacturers take into account in their designs. Compared to plastic packaging, they are significantly more expensive.

Alumina (Al_2O_3) is by far the most common of all ceramics, having been used over the centuries in porcelain and fine dinnerware. Aluminum nitride (AlN) and beryllia (BeO) have superior material properties (e.g., better thermal conductivity), but the latter is very toxic. Aluminum nitride substrates tend to be costly in particular because of required complex processing due to the difficulty of sintering the material.

A ceramic package is made of laminates, each formed and patterned separately, then brought together and cofired (sintered) at an elevated temperature—typically between 1,500°C and 1,600°C (see Figure 8.8). Recent advances have led to low-temperature cofired ceramics (LTCC), such as the Dupont 951 Green Tape™, with sintering temperatures near 800°C. Powders are first mixed together with special additives and extruded under a knife edge to form a thin laminate sheet. This "green" unfired soft tape, approximately 0.1 to 0.3 mm thick, is peeled from the supporting table, then cut and punched using precise machining tools. Patterns of electrical interconnects are screen printed on each sheet using a slurry of tungsten powder or tungsten-molybdenum. This process also fills via holes with metal. Vias left unfilled with tungsten can be later used as fluid- or pressure-access ports through the ceramic. Several "green" sheets are aligned and press laminated together, then cofired at an elevated temperature in a reducing atmosphere to sinter the laminate stack into a monolithic body. A typical integrated circuit package consists of three laminates, but as many as sixteen may be simultaneously cofired, naturally at a higher material cost. An appropriate metal finish is then applied to the tungsten, followed by plating of nickel. If necessary, pins or leads are brazed to the package. The leads are typically made of ASTM F-15 alloy (also known as Kovar®, it is an alloy that consists of 52% iron, 29% nickel, and 18% cobalt) that has a thermal expansion coefficient matched to that of alumina. The brazing material is often a silver-copper eutectic alloy. A final nickel and electroless gold-plating step ensures that wires can be bonded to the leads. A BGA ceramic package has no pins brazed; rather, it has arrays of solder balls connected to electrical feed throughs. One

Table 8.5 Material Properties of Some Notable Ceramics As Compared to Silicon

Ceramic	*Relative Permittivity*	*Thermal Conductivity (W/m•K)*	*Thermal Expansion (10^{-6}/°C)*	*Density (g/cm^{-3})*
Alumina (Al_2O_3)	9.7	40	7.2	4
Aluminum Nitride (AlN)	10	150	2.7	3.2
Beryllia (BeO)	6.8	300	7	2.9
Borosilicate glass	3.7	2	3.2	2.1
Silicon	11.8	157	2.6	2.4

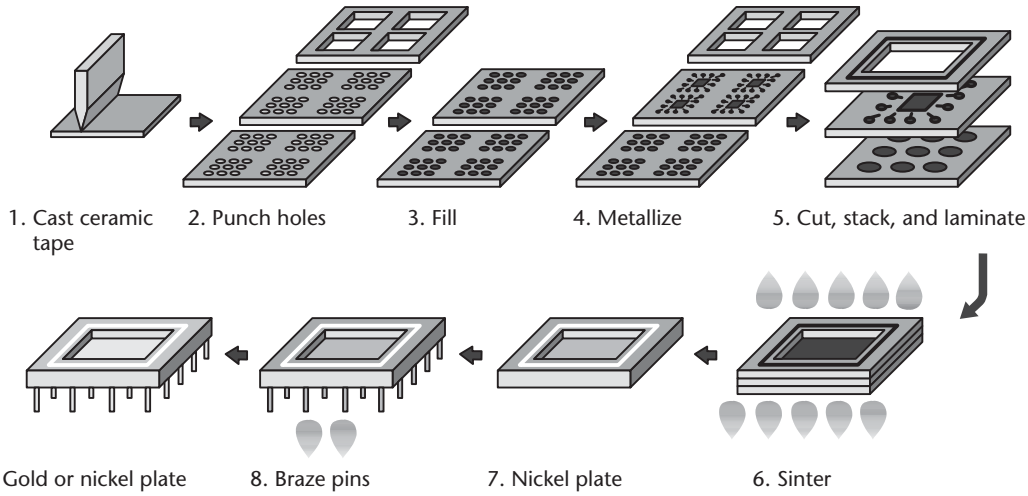


Figure 8.8 Process flow for the fabrication of a cofired laminated ceramic package with electrical pins and access ports. (*Courtesy of:* the Coors Electronic Package Company of Golden, Colorado.)

attractive feature of ceramic is the ability to screen print on its surface a network of thick-film resistors that can be later trimmed with a laser for sensor calibration.

Whether custom or standard, a ceramic package often consists of a base or a header onto which one die or many dice are attached by adhesives or solder. Wire bonding is suitable for electrical interconnects. Flip-chip bonding to a pattern of metal contacts on the ceramic works equally well. The final step after mounting the die on the base and providing suitable electrical interconnects involves capping and sealing the assembly with a lid whose shape and properties are determined by the final application. For instance, the lid must be transparent for optical MEMS or must hermetically seal a vacuum, as is the case for the infrared bolometer from Honeywell or the DMD from Texas Instruments (see Chapter 5). By contrast, a plastic cover provides a cost-effective solution for low-cost devices. For example, disposable blood pressure sensors used for arterial-line measurement in intensive care units are protected by a plastic cover that includes an access opening for pressure [21]. A special gel dispensed inside this opening provides limited protection (particularly against biological solutions and electrical charge) to the device while permitting the transmission of pressure to the sensitive silicon membrane (see Figure 8.9).

Ceramic packaging of optical MEMS can be complex and costly. This is certainly true for DMD packages that have undergone a continuous evolution from their early application in airline ticket boarding (ATB) printers to today's high-resolution display arrays [22]. The DMD type-A package for SVGA displays consists of a 114-pin alumina (Al_2O_3) ceramic header (base) with metallization for electrical interconnects and a Cu-Ag brazed Kovar seal ring (see Figure 8.10). Wire bonds establish electrical connectivity between the die and metal traces on the ceramic header. A transparent window consisting of a polished Corning 7056 glass fused to a stamped gold-nickel-plated Kovar frame covers the assembly. Resistance seam welding of the seal ring on the ceramic base to the Kovar glass frame provides a permanent hermetic seal. Two zeolite getter strips attached to the inside of the glass window ensure long-term desiccation. The particular choice of metal and glass window materials minimizes the mismatch in coefficients of thermal expansion (4×10^{-6} and 5×10^{-6} per

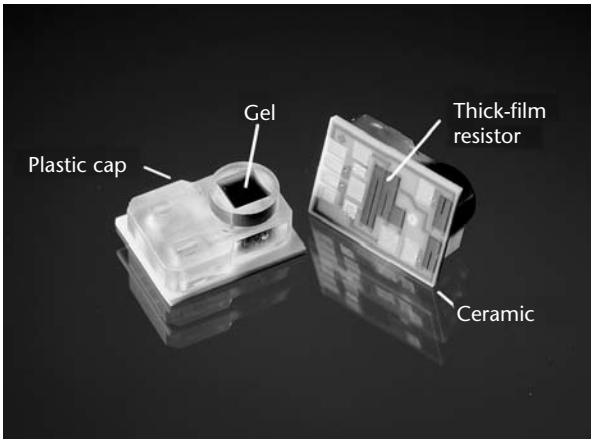


Figure 8.9 Photograph of a disposable blood pressure sensor for arterial-line measurement in intensive care units. The die (not visible) sits on a ceramic substrate and is covered with a plastic cap that includes an access opening for pressure. A special black gel dispensed inside the opening protects the silicon device while permitting the transmission of pressure. (*Courtesy of:* GE NovaSensor, Fremont, California [21].)

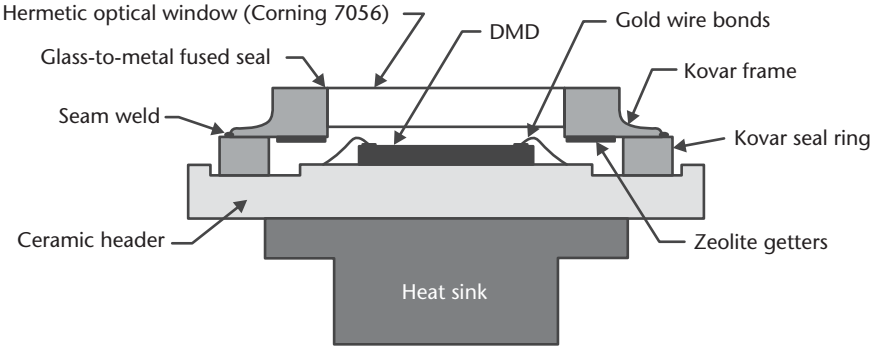


Figure 8.10 Illustration of the DMD type-A ceramic package. The assembly includes a hermetically sealed optical window for high-resolution projection display [22].

degree Celsius for Kovar and Corning 7056, respectively) and reduces stresses during the high temperature (~1,000°C) metal-to-glass fusing process. Antireflective coatings applied to both sides of the glass window reduce reflections to less than 0.5%. A heat sink attached to the backside of the ceramic package by means of adhesives keeps the temperature of the DMD within tolerable limits.

Metal Packaging

In the early days of the integrated circuit industry, the number of transistors on a single chip and the corresponding pin count (number of I/O connections) were few. Metal packages were practical because they were robust and easy to assemble. The standard family of transistor outline (TO)-type packages grew to cover a wide range of shapes, but all accommodated fewer than 10 electrical pins. But the semiconductor industry abandoned the TO packages in favor of plastic and ceramic packaging as the density of transistors grew exponentially and the required pin count increased

correspondingly. Today, TO-type packages remain in use for few applications, in particular high-power discrete devices and high-voltage linear circuits (e.g., operational amplifiers).

Metal packages are attractive to MEMS for the same reasons the integrated circuit industry adopted the technology over 30 years ago. They satisfy the pin count requirements of most MEMS applications; they can be prototyped in small volumes with rather short turnaround periods; and they are hermetic when sealed. But a major drawback is the relatively large expense of metal headers and caps; they cost a few dollars per assembled unit, at least ten times higher than an equivalent plastic package. Early prototypes of the ADXL family of accelerometers from Analog Devices (see Chapter 4) were available in TO-type hermetic metal packages. However, pressure to reduce manufacturing costs has led the company to adopt a standard plastic dual-in-line (DIP) solution and establish first-level packaging (at the die level) using proprietary chip-encapsulation methods.

A metal hermetic package, including the familiar TO-8-type and the tub-like butterfly package, is frequently made of ASTM F-15 alloy (Kovar), though steel is also possible. Because Kovar has a low coefficient of thermal expansion that is matched to fused silica (a common optical material), it is a metal of choice for butterfly packages used in optical and photonic applications. A sheet of metal is first formed into a header or a tub-like housing. Holes are then punched, either through the bottom for plug-in packages or the sides for flat or tub-like packages. An oxide is then grown over the package housing. Metal leads are placed through the holes and beads of borosilicate glass, such as Corning 7052 glass, are placed over the leads. Fusing of the glass to metal at a temperature above the melting temperature of glass (~ 500°C) produces a hermetic metal-to-glass seal. Etching the metal oxides reveals a fresh alloy surface that is then plated with either nickel or gold—both of which allow wire bonding and soldering. Standard headers, butterfly packages, and lids are commercially available and can be readily modified in conventional machine shops. For instance, metal tubes can be brazed to drilled ports in the header and a companion coverlid to provide access to fluids in pressure and flow sensors and microvalves (see Figure 8.11). Similarly, a feed-through tube may be brazed to the sidewalls of a butterfly package for eventual optical interconnecting using a fiber [see Figure 8.7(a)]. In the final packaging assembly, the micromachined structures as well as other components (e.g., optical elements) are mounted directly on the header or within the tub of the package. Wire bonds to the plated package leads establish electrical connectivity. If necessary, optical or fluidic connections are also made, as discussed earlier. Finally, the soldering or seam welding of the header or butterfly package to a coverlid (or cap), most often made of the same alloy, hermetically seals the assembly.

One example of metal packaging applies to the tunable laser from Santur Corp. discussed in Chapter 5. The packaging includes positioning the array of distributed feedback lasers (DFBs), the tilting micromirror, and a host of optical elements within the butterfly package, as well as making the appropriate electrical and fiber interconnections [see Figure 8.12(a)]. The die that holds the micromirror is first attached and wire bonded to a ceramic chip with electrical pads. This micromirror subassembly is mounted on its side over an underlying ceramic plate that also holds the DFB array, two beam splitters, and a InGaAs quadrant detector. The ceramic

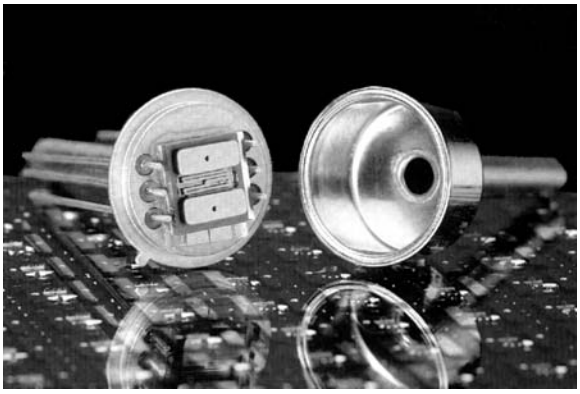


Figure 8.11 Modified by brazing two tubes to the header and the cap, the TO-8 metal can become suitable for packaging fluidic microdevices, such as this microvalve from Redwood Microsystems. (*Courtesy of:* A. Henning, Redwood Microsystems of Menlo Park, California.)

plate sits on a first thermoelectric cooler (TEC) that controls the temperature of the DFBs and performs the fine tuning of the output wavelength. The two beam splitters sample a fraction of the laser light (typically less than 1%) onto the quadrant detector to feed the spatial position of the beam back to the electronics that control the angles of the micromirror.

An etalon and a standard detector epoxied on a second TEC play the role of a built-in wavelength locker [see Figure 8.12(b)]. A fraction of the light sampled by the beam splitters passes through the etalon onto the detector for locking to the ITU grid (see Chapter 5). A second TEC maintains the temperature of the etalon to a predetermined value.

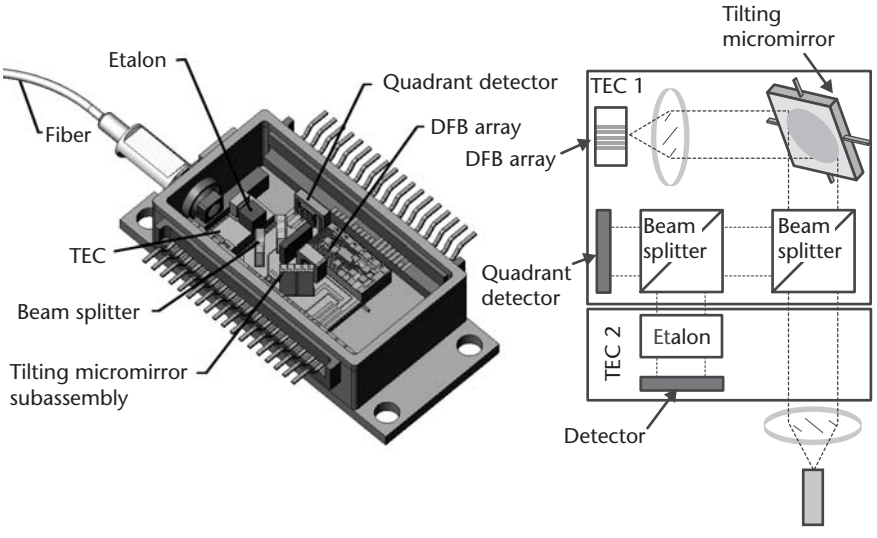


Figure 8.12 (a) A rendering of the packaging for the tunable laser from Santur Corp.; and (b) a block diagram of the components within the packaging of the tunable laser. A first thermoelectric cooler controls the temperature of the DFB array. A second cooler controls the temperature of an etalon and a detector that act as a wavelength locker.

Packaging solutions for harsh environments, namely those found in heavy industries and aerospace, can be complex and costly. The custom requirements of the application, coupled with the lack of high-volume market demand, have turned packaging for harsh environments into a niche art. One particularly interesting design is the metal packaging of media-isolated pressure sensors for operation in heavy-industrial environments. The design immerses the silicon pressure sensor within an oil-filled stainless steel cavity that is sealed with a thin stainless steel diaphragm. The silicon pressure sensor measures pressure transmitted via the steel diaphragm and through the oil. The robust steel package offers hermetic protection of the sensing die and the wire bonds against adverse environmental conditions (see Figure 8.13).

Each stainless steel package is individually machined to produce a cavity. The die is attached to a standard header with glass-fired pins and wire bonded. This header is resistance welded to the stainless-steel package. Arc welding of a stainless-steel diaphragm seals the top side of the assembly. Oil filling of the cavity occurs through a small port at the bottom that is later plugged and sealed by welding a ball.

Molded Plastic Packaging

Unlike metal or ceramic packages, molded plastic packages are not hermetic. Yet they dominate in the packaging of integrated circuits because they are cost-effective solutions (costing on average a few pennies or less per electrical pin). Advances in plastic packaging have further improved reliability to high levels. Today's failure rates in plastic-packaged logic and linear integrated circuits are less than one failure in every ten billion hours of operation [23].

There are two general approaches to plastic packaging: post molding and pre-molding (see Figure 8.14). In the first approach, the plastic housing is molded after the die is attached to a *lead frame* (a supporting metal sheet). The process subjects the die and the wire bonds to the harsh molding environment. In premolding, the die is attached to a lead frame over which plastic was previously molded. It is attractive

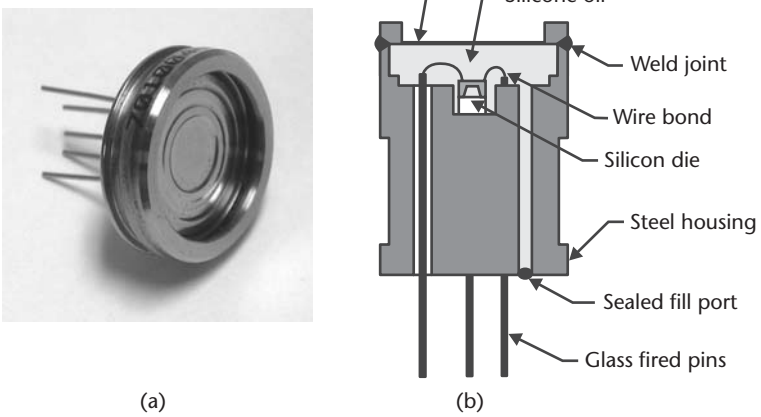


Figure 8.13 (a) Photograph, and (b) cross-sectional schematic of a pressure sensor mounted inside an oil-filled, stainless-steel package. Pressure is transmitted via the stainless-steel diaphragm and through the oil to the silicon sensor. (*Courtesy of:* GE NovaSensor of Fremont, California.)

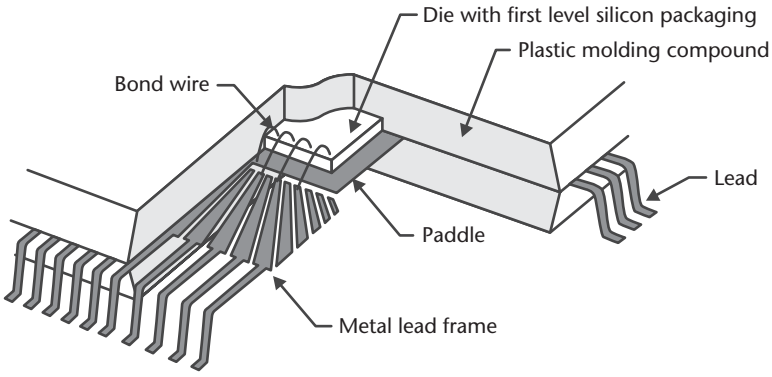


Figure 8.14 Schematic showing a sectional view of a post-molded plastic package. The die is first mounted on a center platform (the paddle) and wires bonded to adjacent electrical leads. The paddle and the leads form a metal *lead frame,* over which the plastic is molded. A MEMS die should include a first level of packaging (e.g., a bonded silicon cap) as protection against the harsh effects of the molding process. This particular illustration is of a plastic quad-flat pack (QFP) with electrical leads along its entire outer periphery.

in situations where the risk of damaging the die is high or if openings through the plastic are necessary (e.g., for pressure or flow sensors). However, it tends to be more expensive than post molding.

The metal lead frame in either approach is an etched or stamped metal sheet consisting of a central platform (paddle) and metal leads supported by an outer frame. The leads provide electrical connectivity and emanate from the paddle in the shape of a fan. The metal is typically a copper alloy or Alloy-42 ($Ni_{42}Fe_{58}$); the latter has a coefficient of thermal expansion 4.3×10^{-6} per degree Celsius that matches that of silicon.

In postmolded plastic packaging, the lead frame is spot-plated with gold or silver on the paddle and the lead tips to improve wire bonding. The die is then attached with adhesive or eutectic solder. Wires are bonded between the die and the lead tips. Plastic molding encapsulates the die and lead frame assembly but leaves the outer edges of the leads exposed. These leads are later plated with tin or tin-lead to improve wetting during soldering to printed circuit boards. Finally, the outer frame is broken off and the leads are formed into a final S-shape (see Figure 8.14).

The sequence of process steps differs for premolded plastic packages. First, a plastic body is molded onto a metal lead frame. The molded thermosetting plastic polymer encapsulates the entire lead frame with the exception of the paddle and the outer edges of the leads. Deflashing of the package removes any undesirable or residual plastic on the die bonding areas. The molded body may contain ports or openings that may be later used to admit a fluid (e.g., for pressure or flow sensing). The lead frame is spot-plated with gold or silver to improve wire bonding and soldering. At this point, the die is attached and wire bonded to the lead frame. A protective encapsulant, such as RTV or silicone gel, is then dispensed over the die and wire bonds. Finally, a premolded plastic cap is attached using an adhesive or ultrasonic welding. If necessary, the cap itself may also contain a fluid access port (Figures 8.15 and 8.16).

The molding process is a harsh process involving mixing the component for the thermosetting plastic at approximately 175°C, then flowing it under relatively high

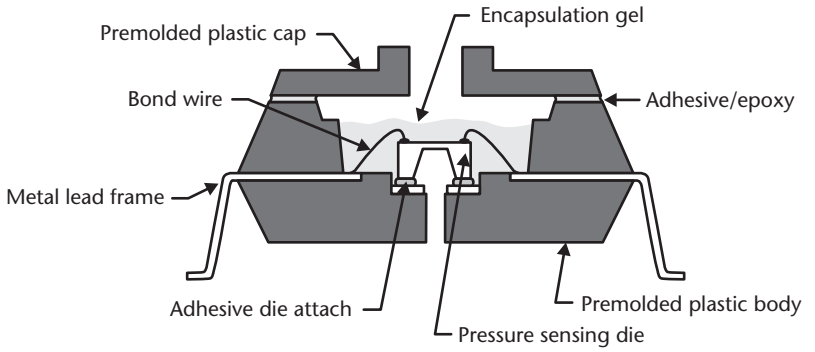


Figure 8.15 Illustration of a premolded plastic package [24]. Adapting it to pressure sensors involves incorporating fluid ports in the premolded plastic housing and the cap.

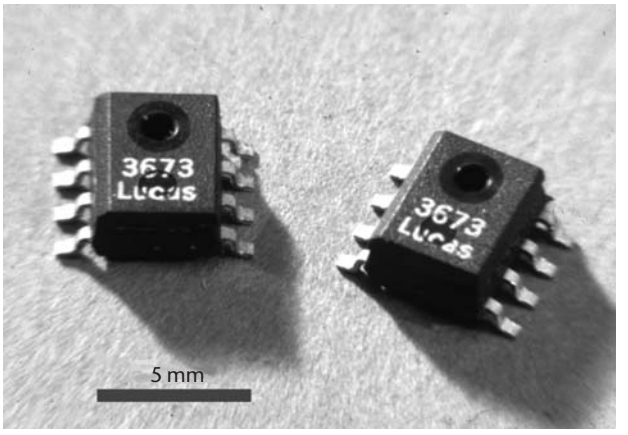


Figure 8.16 Photograph of the NovaSensor NPP-301, a premolded plastic, surface mount (SOIC-type) and absolute pressure sensor. (*Courtesy of:* GE NovaSensor of Fremont, California.)

pressure (~ 6 MPa) into the mold cavity before it is allowed to cool. The plastic material is frequently an epoxy. Novolac epoxies are preferred because of their improved resistance to heat. The temperature cycle gives rise to severe thermal stresses due to the mismatch in coefficients of thermal expansion between the plastic, the lead frame, and the die. These stresses may damage the die or cause localized delamination of the plastic. The material properties of the plastic, and especially its coefficient of thermal expansion, are carefully adjusted by the introduction of additives to the epoxy. Fillers such as glass, silica, or alumina powder make up 65% to 70% of the weight of the final product and help tailor its coefficient of thermal expansion as well as its thermal conductivity. In addition, mold release agents (e.g., synthetic or natural wax) are introduced to promote releasing the plastic part from the mold. Flame-retardant materials, typically brominated epoxy or antimony trioxide, are also added to meet industry flammability standards. Carbon and other organic dyes give the plastic its all-too-common black appearance, which is necessary for laser marking.

Plastic packaging for integrated circuits are governed by standards set forth by the Electronics Industries Association (EIA), the Joint Electron Device Engineering Council (JEDEC), and the Electronics Industry Association of Japan (EIAJ) (see Table 8.6). While plastic packaging for MEMS is not governed by any standards yet, it often uses standard or slightly modified integrated-circuit plastic packages. The development of new plastic packaging technologies for MEMS will likely remain in the far future because of the prohibitive associated costs.

Quality Control, Reliability, and Failure Analysis

When questioned about the reliability of a MEMS or micromachined component, the spontaneous reaction of an average consumer is often negative, vaguely pointing that these devices just cannot be "reliable." Myth more than scientific reality influences the minds of people in developing such an opinion. For example, there is a perception that small size cannot instill a sense of reliability. Yet, it is the small dimensions that generally increase immunity to shocks, make friction miniscule, and reduce electrical power consumption and heat dissipation. Only when reminded that most automobiles in the world depend on micromachined sensors for engine operation and passenger safety does the negative image in the individual's mind begin to change. As the MEMS industry continues to mature, it will further improve its existing quality and reliability procedures; as products permeate through society, the consumer will become more at ease with the reliability of these tiny components, making them one day synonymous with that of a sister industry—electronic integrated circuits.

Table 8.6 Selected Standard Molded Plastic Packages for Integrated Circuits*

	Type	*Pin Count*	*Description*
Surface Mount	Small outline transistor (SOT)	Min. 3, max. 8	Small package with leads on two sides
	Small outline IC (SOIC)	Min. 8, max. 28	Small package with leads on two sides
	Thin small outline package (TSOP)	Min. 26, max. 70	Thin version of the SOIC
	Small outline J-lead (SOJ)	Min. 24, max. 32	Same as SOIC but with leads bent in J shape
	Plastic leaded chip carrier (PLCC)	Min. 18, max. 84	J-shaped leads on four sides
	Thin QFP (TQFP)	Min. 32, max. 256	Wide but thin package with leads on four sides
Through-Hole Mount	Transistor outline 220 (TO220)	Min. 3, max. 7	One in-line row of leads, with heat sink
	Dual in-line (DIP)	Min. 8, max. 64	Two in-line rows of leads
	Single in line (SIP)	Min. 11, max. 40	One in-line row of leads
	Zigzag in line (ZIP)	Min. 16, max. 40	Two rows with staggered leads
	Quad in line package (QUIP)	Min. 16, max. 64	Four in-line rows of leads; leads are staggered

(*Source:* [25].)

*Surface mount devices are generally thinner than through-hole mount packages and accommodate a smaller spacing between adjacent leads (pins).

Quality Control and Reliability Standards

There are no standards that specifically govern the reliability of micromachined components or MEMS in general. Instead, the MEMS industry derives its quality and reliability guidelines from the quality standards of the systems into which MEMS and microsystems are ultimately inserted. For example, the fabrication of micromachined sensors used in automotive applications is frequently subject to the quality management principles of QS 9000 standards initially set forth in 1994 by Daimler-Chrysler, Ford, and General Motors for the entire auto industry in the United States. The QS 9000 standard is itself an evolution of another quality management standard, ISO 9001, which was established by the International Organization for Standardization of Geneva, Switzerland. Similarly, Telcordia® Technologies manages a set of quality and reliability standards specific to products and equipment for the telecommunications industry. Table 8.7 lists a number of standards widely used in industries and applications that have adopted MEMS and microsystems technology.

These standards differ vastly in their impact on the manufacturers of micromachined components. The ISO 9000 series and QS 9000 standards address quality management principles such as methods for process control, documentation, and uniform procedures, but they do not specify particular tests or reliability requirements. These standards leave the details of the qualification tests and reliability specifications to the manufacturer, as long as they are well documented and follow predescribed quality management principles. The typical result of the ISO and QS standards is a manufacturing operation with clear controls over its design and manufacturing processes. Mature companies often seek certification by third parties specialized in auditing and reviewing quality management systems.

Unlike the ISO 9000 and QS 9000 standards, the Telcordia, IEEE, and MIL standards detail specific environmental and operational tests for qualification and reliability. These tests have two purposes. The first one is to evaluate the product's performance under rigorous environmental conditions, in particular, shock and vibration, temperature, humidity, and occasionally salt spray and altitude. Shock and vibration tests simulate situations observed during handling and shipping, or in high-vibration environments such as portable applications. Temperature testing validates the overall thermal design of the product. Humidity tests checks for condensation effects on performance and reliability, particularly as they affect corrosion. Salt spray (as specified in MIL-STD-810) is largely unique to marine or military applications and is not common for most commercial applications. Altitude testing is useful for evaluating high-voltage insulation because low pressure induces chemical changes in the insulating material.

The second purpose is to precipitate a failure of the product by stressing it under the effect of carefully applied operational and environmental conditions over extended durations. These tests include humidity and temperature cycling, thermal shock, operation in damp heat, mechanical stressing, and burn-in—collectively, they form the basis of accelerated life testing and they are casually referred to as "shake and bake." Burn-in, specified under MIL-STD-883, is common in the reliability testing of electronic components and seeks to detect latent defects that will result in infant mortality (failures that occur at a very early stage). During burn-in,

Table 8.7 A Select List of Key Reliability and Quality Standards for Systems and Applications that Are Likely to Incorporate MEMS or Microsystems

Standards	*Description*	*Organization/Regulatory Body*	*Web Site*
ISO 9000 series	Principles for general quality management	International Organization for Standardization	www.iso.ch
QS 9000	Automotive quality management	Automotive Industry Action Group	www.aiag.org
IEEE 1332	Program for reliability of electronic systems	IEEE	www.ieee.org
IEEE 1413	Methodology for reliability prediction	IEEE	
MIL-HDB-217	Reliability prediction for electronics	U.S. Department of Defense	dodssp.daps.mil
MIL-STD-202	Test methods for electronic components	U.S. Department of Defense	
MIL-STD-883	Test methods for microelectronics	U.S. Department of Defense	
GR-63-CORE	Standard for environmental criteria for telecom equipment	Telcordia Technologies	www.telcordia.com
GR-463-CORE	Standard for the reliability of optoelectronic devices	Telcordia Technologies	
GR-1209-CORE	Standard for the reliability of branched optical devices	Telcordia Technologies	
GR-1221-CORE	Standard for the reliability of passive optical devices	Telcordia Technologies	
21 CFR Parts 800-1299	Clearance pursuant to Title 21 Code of Federal Regulations	U.S. Food and Drug Administration, Center for Devices and Radiological Health	www.fda.gov/cdrh

temperature stresses and electrical voltages are applied for an extended duration of time with operation at maximum load under different ambient temperatures.

One example of extensive reliability standards is the GR-CORE series from Telcordia Technologies for telecommunication components and equipment. These standards clearly outline test conditions for shock, vibration, temperature, and humidity cycling, accelerated aging as well as other test parameters to evaluate the rate of infant mortality and gauge the long-term reliability of the product. Many of the tests defined under the Telcordia standards originate from the MIL standards defined by the U.S. Department of Defense for the operation of components for military applications. For example, the GR-63, 463, 1209, and 1221 CORE standards that define the reliability tests specifically for optoelectronic and passive fiber-optical components (see Table 8.8) explicitly reference the MIL-STD-202 and 883 standards. Procedures and methods also accompany the MIL standards to guide the user in performing the tests and interpreting the data. For example, the MIL-HDBK-H-108 provides sampling procedures and tables for reliability and life testing, and the MIL-HDBK-217 discusses reliability prediction (e.g., calculation of failure rates and mean time to failures) of electronic equipment [26].

An industry of professional consultants and advisors specializing in quality and reliability standards has come to exist. The manufacturer of MEMS products is well

Table 8.8 A Summary of the Key Reliability Tests Specified Under the Telcordia Standards GR-63/463 and GR-1209/1221 for the Qualification of Devices for Optical Telecommunications

Test	*GR-63/463-CORE Reliability Assurance for Optoelectronic Devices*	*GR-1209/1221-CORE Reliability Assurance for Branched and Passive Fiber-Optic Devices*
Mechanical shock	500G for 1 ms, 5 times/axis	500G for 1 ms, 2 times/axis; 200G for 1.33 ms, 2 times/axis
Nonoperational vibration	20G, 20–2,000 Hz, 4 min/cycle, 4 cycles	20G, 20–2,000 Hz, 4 min/cycle, 4 cycles
Operational vibration	5.0G, 10–100 Hz; 2.4G, 100–200 Hz	10–55 Hz, 1.52 mm amplitude, 20 min per 3 axes
Thermal shock (air-to-air)	15 cycles, 0° to 100°C	—
Solderability	+260°C for 10s	—
Accelerated aging (operational)	70°C or 85°C, > 2,000 hours	—
High-temperature storage	+85°C, 2,000 hours	+85°C, RH<40% RH, 2,000 hours
Low-temperature storage	–40°C, 2,000 hours	–40°C, 2,000 hours
Temperature cycling	–40°C to +70°C, >100 cycles	–40°C to +70°C, >100 cycles; –40°C to +70°C, 10% to 80% RH, 42 cycles
Damp heat	+85°C/85% RH, 1,000 hours	+85°C/85% RH, 500 hours
Internal moisture	<5,000 ppm water vapor	—
ESD threshold	±500-V discharge, each pin set	—
Fiber pull	1.0 kg, 3 times, 5-s duration	—
Fiber twist and flex tests	—	0.5-kg load, 100 cycles
Side pull	—	0.25 to 0.5 kg-load at 90° angle
Cable retention	—	0.5 to 1 kg-load for 1 minute
Water immersion	—	43°C, pH 5.5, for 336 hours

(*Source:* [27].)

advised to seek such professional recourse. The user of MEMS products will often demand that those products are certified under one or many quality standards that are most applicable to the user's industry. However bureaucratic these standards may on occasion be perceived by the general scientific community, they are of paramount importance to the MEMS industry as it transitions from prototyping experimentation to mature manufacturing.

Statistical Methods in Reliability

If one defines reliability as the *probability* that a device will perform its specified functions without failing over an expected *operating time* within defined operating and environmental conditions, then it becomes clear that statistics play an important role in assessing and predicting the reliability of a product. This section introduces a few key concepts and terms commonly used in the theory of reliability. The reader is referred to the books by Băjenescu et al. [28] and Kececioglu [29, 30] for further insight on the methodologies of reliability.

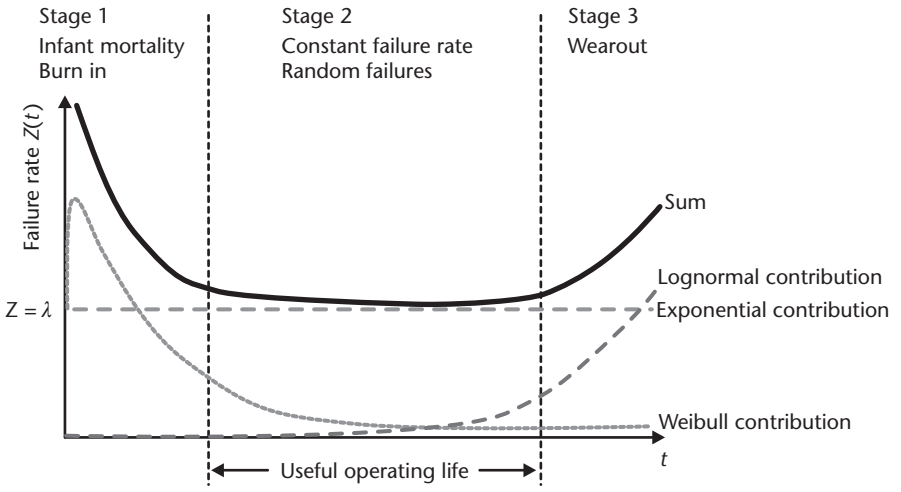
Failure is defined as the termination of the ability of a product to meet required specifications or perform a required function. Failures are random events that are statistically independent and can thus be described by standard probability distribution functions that follow the Poisson process. Depending on the underlying physics,

the random failure events can have different probability distribution laws (e.g., exponential, normal, lognormal, Weibull, gamma, and Rayleigh [29, 31]). The operating time is a duration for which the product performs its required function. For a nonrepairable product, the mean operating time is also referred to as the mean time *to* failure (MTTF). For a product that can be completely repaired, the mean time of operation becomes the mean time *between* failures (MTBF). As most electronic and micromachined components are often difficult to repair after failure, we will limit the discussion of lifetime to MTTF.

Knowledge of the probability distribution function, $f(t)$, is necessary to compute the probability of a unit failing as well as the failure rate and MTTF [28, 29]. The probability of a failure at time t, defined as $F(t)$, is the area under the distribution function, mathematically given by the integral over a time period t. It is mostly a mathematical concept that is not widely used in specifying product reliability. Instead, failure rate and MTTF are the two key and practical specifications in the assessment and prediction of reliability. The failure rate, also known as hazard rate, $Z(t)$, is a measure of the instantaneous speed of failure, effectively the number of failures over a given period of time. Consequently, it has units of failures per unit time, most commonly one failure in one billion hours (10^{-9}/hr) also known as failure in time (FIT). Mathematically, it can be shown that $Z(t) = f(t)/[1-F(t)]$. Experimentally, the failure rate is calculated as the ratio of the observed number of failures occurring in a time interval to the number of functional devices at the beginning of this time interval, normalized to the length of the time interval [28]. The larger the number of devices and the longer the observation time are, the higher the statistical confidence becomes in the measured failure rate. This confidence is mathematically reflected by multiplying the measured failure rate by the statistical chi squared (χ^2) parameter [31]. When the observation time is impractically long to achieve reasonable confidence, temperature-based accelerated life testing (described later) becomes an invaluable tool to extrapolate values for the failure rate and MTTF.

The experimentally observed failure rate of many high-technology products, including electronic, fiber-optical, and micromachined components, exhibits a characteristic time-dependent behavior that is best described by the "bath tub" curve (see Figure 8.17). This curve shows an early stage in the life of the product with a rapidly decreasing failure rate resulting from better screening, improving reliability, and lower infant mortality. A second stage characterized by a rather constant failure rate defines the mean useful life of the component in the field. A rising failure rate brought by an increase in wear signals the onset of the last stage and the end of the useful life.

Reliability scientists model the bath-tub curve as a superposition of three different probability distribution functions, one for each stage in the curve. The Weibull distribution function best models the early stage, whereas the lognormal distribution is used to model the third stage. The exponential distribution is best to describe the middle span because it models a constant failure rate that we denote as λ. The overall failure rate curve is the sum of all three contributions (see Figure 8.17). The middle span is one that attracts most attention, as it describes the reliability of the product during its most useful life. A key characteristic of the exponential distribution function is its time-independent failure rate, which allows for varying the combination of the number of devices under test and the hours of testing. For



$$\lambda = \frac{\chi^2(2n+2)\cdot 10^9}{2\cdot N\cdot T}$$

λ = Failure rate in FIT
n = Number of observed failures
N = Number of functional devices at the beginning of period
T = Duration of observation period
$\chi^2(2n+2)$ = Statistical chi squared parameter

Figure 8.17 The reliability bath-tub relationship between failure rate $Z(t)$ and time t. It consists of three temporal stages, each with its listed characteristics. The failure rate in the middle span of the curve is time independent and equal to λ. The overall failure rate can be modeled as the sum of the contributions of three probability distribution functions. Using the exponential distribution function suited only for the middle span, one can calculate the MTTF to equal $1/\lambda$.

example, if 10,000 unit hours of testing is required, then one can test 10 units for 1,000 hours, or 100 units for 100 hours or some other combination. The constant failure rate (λ) can then be expressed in failures per unit of time. For an exponential distribution, one can mathematically show that the MTTF is equal to $1/\lambda$ [28]. Clearly, the exponential approximation is valid only for the middle span of the curve and should not be used elsewhere.

Accelerated Life Modeling

An accelerated life model is one that predicts failure as a function of applied operating and environmental stresses. Shock and vibration, temperature and humidity cycling, mechanical stress, and burn-in belong to a category of qualitative accelerated life testing intended to bring out failure modes that would normally manifest themselves in later stages of the product's life. Once a failure is observed, appropriate corrective actions are taken to eliminate the origin of the failure. By contrast, another category of accelerated life testing is quantitative in nature and aims to predict a failure rate and an MTTF. Stress tests such as operation in high heat, high humidity, and high voltages are good examples. These tests rely on the theory of rate processes [30], which is generally described by an exponential dependence on the stress parameter to determine the degradation in a particular life characteristic due to the applied stress—this dependence is known as the acceleration factor. The Eyring equation is a generalized model that can take into account multiple stress

parameters, including temperature, humidity, and voltage [32]. The Arrhenius equation, a special case of the Eyring equation, is a well-known example of a rate process where the stress parameter is only temperature. If the failure rate is constant in time and the exponential distribution function is applicable, then the degrading life characteristic is time to failure (lifetime) and the corresponding acceleration factor is proportional to $\exp(-E_a/kT)$ where E_a is the activation energy, k is the Boltzmann constant, and T is temperature [33]. Should there be an indication that the failure rate is not constant in time, then a more appropriate probability distribution function must be used, resulting in a different degrading life characteristic and a different expression for the Arrhenius acceleration factor [33, 34]. The Arrhenius equation is very useful to model failures that depend on chemical reactions, diffusion processes, and migration processes. This includes failure modes in die attach, epoxies, solder, metal interconnects, thin films, and semiconductor junctions. The Arrhenius model has a limitation specific to micromachined components and MEMS: it is not suitable to analyze accelerated failures resulting from mechanical fatigue, a phenomenon that has been observed in polycrystalline and amorphous materials used in the fabrication of MEMS. This limitation is of most significance to surface-micromachined actuators made of polysilicon or metal alloys.

To find the activation energy, the time to failure is measured at a few elevated temperatures. It is advantageous to make the measurements at the highest possible temperatures in order to shorten the observation time, provided that the applied temperatures do not alter the nature of the failure or damage the device under test. For example, it is not possible to apply a temperature that exceeds the flow temperature of epoxies or solder because the physics of the failure modes will certainly change and the accelerated life model will fail. An exponential curve fit is then applied to the measured data. The slope of the logarithm of the time to failure plotted against the inverse of absolute temperature (in Kelvins) is equal to the activation energy. The MTTF or lifetime at the normal operating temperature (often room temperature) is extrapolated using the Arrhenius equation (see Figure 8.18).

Major Failure Modes

It is evident from the diversity of materials, fabrication processes, and products introduced in the earlier chapters that the possible failure modes would be numerous and equally diverse. The purpose of this section is not to replace standard failure mode and effect analysis (FMEA) methodology to unravel the details of a failure, but rather to point to a few common failure modes that the industry has learned to address.

Decades of development and millions of deployed units have provided plenty of insight and knowledge into the reliability of micromachined electromechanical sensors, in particular pressure sensors and accelerometers. These products have evolved through multiple generations and can now operate and survive under extreme environmental conditions. Over the years, engineers incorporated many design and manufacturing improvements, each addressing one or more possible failure modes. In some instances, these details have become public knowledge. For example, rounding of the corners is now a common practice to reduce stress concentration in micromechanical structures. But in many other instances, manufacturers consider these details as trade secrets, especially when utility patents cannot be

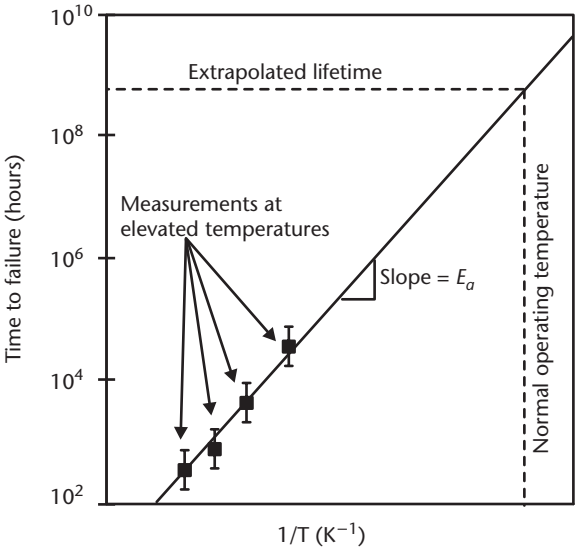


Figure 8.18 The Arrhenius model is a useful tool in accelerated life testing to extrapolate the lifetime of a device at normal operating temperatures. Measurements of time to failure are made at a few elevated temperatures, and an exponential curve fit is applied to the data to calculate the activation energy. This extrapolation method assumes a constant failure rate and thus is specific to the exponential distribution function that is graphically represented by the scatter in the measured data at any particular temperature.

obtained or are difficult to enforce. This level of secrecy becomes even more entrenched in packaging and assembly.

Much work remains to be completed before the reliability of relatively new product entries, especially those with integrated optical, fluidic, RF, or acoustic functions, reaches a level similar to that of widely deployed micromachined sensors. Clearly, these new products have benefited from the existing body of knowledge on reliability, but there remain failures specific to them. For instance, the reliability of the junction between silicon (or glass) and a fluidic interconnect is constantly subject to improvement. In another example, it is not clear whether outgassing from epoxies or other adhesives used in the packaging of optical elements affects the reflectivity of micromachined mirrors or interferes with the high voltages typically used in the electrostatic actuators that drive those mirrors.

Failure modes stem either from weaknesses in the design itself (of both the micromachined device and packaging) or from process variations that result in critical departures from the nominal design. Addressing the former usually follows a systematic course of tests and simulations to pinpoint the exact origin of the design weakness. For example, computer-aided simulations are very useful in identifying nodes of high stress that can result in fracture. Addressing the weaknesses stemming from process control is a more tedious and time-consuming task, which necessitates patience, experience, technical flair, and good organization. For example, the source of an occasional electrical short or open in an electrostatic actuator may be quite difficult to diagnose. It may be the result of a variation in the thickness of a metal trace over a topographical step, or it may be caused by defects in an insulating layer, or there could be yet another plausible explanation. Countless companies proud of

their new and innovative MEMS product ideas had to face these types of reliability questions as they transitioned from a prototyping phase to a manufacturing phase. Analog Devices, Inc., invested tremendous time and resources to resolve the stiction problem that plagued their early accelerometer designs. The following are major observed failure modes that span both design and process control.

Cracks and Fractures

Cracks can occur in a number of locations in a microstructure and are the result of a large stress that exceeds the fracture stress of the material or fatigue [35]. Observing and diagnosing a large fracture is readily achieved under an optical or scanning electron microscope. Acoustic imaging is occasionally used, but its utility is limited to detecting large embedded defects. However, hairline fractures can seldom be seen. Instead, their existence is indirectly detected by measuring their effect on a number of other parameters (e.g., by looking for anomalies in the frequency response of a mechanical element). Often, a mechanical shock is the causing event of fracture [36]. Naturally, the mass of the micromechanical structure must be relatively large for the shock to pose any real risk. For instance, a 10-μm thick, 1 mm^2 membrane in a pressure sensor has very little mass (24 μg) and can sustain shocks up to 100,000G [37]!

A mechanical shock can originate from several sources. It can be intentional, as is the case for accelerometers or during the reliability testing of the product, or accidental (e.g., from the saw during wafer dicing or rough handling during packaging). A mechanical shock has a typical duration of tens to hundreds of microseconds, which greatly exceeds the acoustic transit time defined as the time it takes an acoustic wave to travel through the longest dimension of the micromechanical structure (typically on the order of a few microseconds). Therefore, the micromechanical structure behaves as if the shock is static and wave analysis is not necessary [38]. This is not the case with ultrasonic pulses. Such short pulses (of duration $\leq$1 μs) are common in ultrasonic cleaning, wire bonding during packaging, or in micromachined resonators that specifically utilize ultrasound in their operation. Waves will reflect back at geometrical discontinuities and material boundaries due to mismatches in acoustic impedances, causing standing waves and local amplification of stresses possibly beyond the fracture stress. In such situations, a complete wave analysis using computer-aided modeling is useful to identify the fragile boundaries.

Under shock, the displacement of a freestanding micromechanical structure may exceed its maximum allowed design limit, thus causing excessive stress at one or more particular locations (often corners) and consequently failure. Good mechanical designs take such shocks into account. For example, some well-designed accelerometers have travel limiters for the inertial mass (see Chapter 4) in order to minimize the stress on the supporting spring in the presence of a large shock. Sharp corners are also responsible for fracture under even small shocks because they concentrate mechanical stress. Designers have learned to round the profile of corners to reduce this risk of failure. For example, the corner formed by the intersection of the {111} planes of the thick frame and the thin membrane of a pressure sensor (see Chapter 4) is characteristically sharp, and virtually all manufacturers of modern commercial pressure sensors perform some type of corner rounding. While knowledge about such failure modes is now common, details of the corresponding

solutions remain proprietary to the manufacturers. For instance, a dilute silicon etchant will round the corner mentioned earlier, but the duration of the etch, the type of etchant used, and concentration are all considered trade secrets.

Shocks can also cause fracture by exciting undamped mechanical resonant modes. By virtue of the instantaneous energy they impart on an object, shocks have a very broad spectral signature (up to hundreds of kilohertz) that overlaps with the resonant frequencies of many micromachined elements. Proper mechanical design should address the appropriate damping of any resonant mode that may be excited by shock. This is of particular significance to suspended structures made of single-crystal silicon packaged in vacuum because they make excellent low-frequency resonators with high quality factors. While single-crystal silicon is generally less susceptible to fracture than polysilicon or amorphous silicon, its crystalline nature offers little internal damping. External damping (e.g., air viscous damping) is the only means to reduce the quality factor of single-crystal-silicon beams and the risk of fracture under shock.

Corner rounding, travel limiters, and damping are examples of design modifications intended to improve immunity to shock. The cause-and-effect relationship tends to be well understood either through modeling or extensive testing. Yet, there are other factors related to fabrication and process control that can have an equally dramatic effect on immunity to shock but whose details can be quite arduous to unravel. For instance, a poorly controlled electrochemical etch in the fabrication of a pressure sensor can yield a membrane lacking thickness uniformity that can jeopardize its mechanical properties, including sensitivity to mechanical shock. Similarly, increased scalloping and undercut in a poorly controlled deep reactive ion etch (DRIE) can seriously degrade the mechanical properties of a micromachined structure. Investigating a failure that may be connected to poor process control is a costly and time-consuming proposition. Instead, it is more important and economical to implement preventive process controls over the critical fabrication and packaging steps.

Packaging plays a very important role in shielding a micromechanical structure from the effects of externally applied shocks. For example, the elastic nature of most room-temperature vulcanizing (RTV) rubbers used in die attach is useful to minimize the transmission of stress from the package to the sensitive micromechanical elements in a shock [37]. Intermediate substrates and supporting material between the die and the outer casing of the package can also be useful in protection and lessening the impact of a shock.

Delamination

Though not a frequent failure mechanism, delamination is a concern in all multilayered stacks. The fabrication methods of such stacks are detailed in Chapter 3 and include silicon fusion bonding, LIGA, as well as surface micromachining. Delamination is often the result of poor process control. In silicon fusion bonding, an unreliable bond can usually be traced to minute surface defects, chemical contamination, excessive bowing of the substrates under stress, or poor hydration of the surfaces prior to bringing the two wafers into contact. A submicron defect is sufficient to create an imperceptible void between the two surfaces that can later propagate under the effect of aging and other environmental factors, such as shock. Silicon

wafers that were ground and polished may exhibit large intrinsic stresses and excessive bowing, which weaken the bond. Experience has shown that the bond will ultimately fail unless the stress is not annealed at an elevated temperature (>500°C). Stresses play an equally important role in the failure of thin-film stacks, especially those incorporating platinum, palladium, and gold. In such cases, it is standard practice to include an intermediate layer (often chromium or a titanium alloy) to promote better adhesion. A primitive yet effective test for gauging sensitivity to delamination is the *tape peel test:* an adhesive tape is attached to the layers or substrates in question and manually peeled back.

In some circumstances, the area of adhesion between two layers or wafers may be too small to sustain large forces in the event of a shock. This is the case with polysilicon hinges that are anchored to the substrate using small polysilicon staples (see Figure 4.6). A reliability analysis at the design stage should uncover such a weakness and increase the bonding or adhesion area.

Stiction

Stiction is the failure mode that describes the situation when surface adhesion forces are larger than the mechanical restoring force of a suspended micromechanical element. Surface adhesion forces include capillary forces, electrostatic attraction, and van der Waals forces.

Stiction is a serious problem in surface-micromachined devices that occurs immediately after removing the die from the aqueous solution used to etch the sacrificial layer. A liquid meniscus formed on hydrophilic surfaces inside the suspension gap pulls the microstructure towards the substrate, causing the two surfaces to come into contact and stick (see Figure 3.20). Stiction can also occur after deployment in the field if water condenses inside the gap. The humidity tests outlined in the Telcordia and MIL specifications are useful in accelerating failures due to stiction. There are currently two leading methods that can alleviate the impact of stiction induced by liquid capillary forces, both described in great detail in Chapter 3: the use of antistiction coatings and supercritical drying when pulling the structures out of liquid solutions. Coatings remain the topic of much-needed research.

A different type of sticking happens in electrical relays and RF capacitive switches. In relays operating at dc currents with ohmic contact between the poles [39], repeated impacting between the two metal surfaces of the poles degrades the contacts, ultimately leading to stiction. This process is akin to microwelding, due to localized heating in the contact area. In RF capacitive switches, electrical charging within the dielectric layers permanently pulls the switch membrane to the substrate. The electrical charges are normally trapped within the dielectric with no available conduction path. The origin of the charging is not well known, though tunneling into the dielectric is a leading hypothesis [40].

Electrical and Thermal Failure Modes

MEMS share with integrated circuits many electrical and thermal failure modes, such as burnout of electrical interconnects and diode junctions. Metal traces on the silicon die are subject to the same rules of electrical reliability found in the integrated circuit industry, including electrical crowding, electromigration, current

density limitation (typically <100 kA/cm^2), and localized resistive heating in the wiring. Metal patterns over rising topographical features are particularly susceptible to failure because of the potentially thin metal at the corners. Failure of electrical wiring is not limited to on-chip metal traces. Wire bonds are also subject to failure if they are not matched to the current requirements of the device. Additionally, wire bonds can fail under mechanical shock and vibration because they are often made of gold (a dense material) and have long lengths exceeding many millimeters.

Thermal failures occur when there is an excessive temperature rise due to localized heating (e.g., from a high electrical resistance or an accidental current surge) or when there is poor conduction of heat generated within the device (e.g., in thermal actuators that can dissipate several watts). In all cases, proper thermal management at the die and package level is important to mitigate the risk of thermal failures.

A Reliability Case Study: The DMD

The DMD, described in detail in Chapter 5, is an excellent example of a complex microsystem that merges electronic, mechanical, optical, and chemical attributes, thus making its reliability a highly interactive relationship between many diverse operational factors and environmental parameters. The historical evolution of the DMD and its reliability over nearly two decades highlights the depth and breadth of the development effort that has yielded this commercial success.

Early in the 1990s, the lifetime of the DMD was only 100 hours at 65°C, whereas the target application (primarily printing) required a minimum of 5,000 hours. Some parts worked well, but others did not. The design was marginal, and the fabrication processes were not under control. The origins of the failures were largely unknown, which made improving the reliability a daunting task. Texas Instruments undertook a program of extensive testing, characterization, and analysis of the failure modes that gradually increased the understanding of the underlying physics and resulted in new designs that were more robust and reliable.

With the novelty of the DMD design and the emerging nature of the MEMS industry, Texas Instruments had to develop many specialized tests and build the corresponding equipment in house. These tests varied many operational parameters, including temperature, voltage and timing waveform, the number of mirror landings, mirror duty cycle, and light intensity. It then sought to identify statistical relationships that would lead to the location of design parameters that yielded a more reliable performance. They also performed a number of environmental tests, many similar to those defined under the Telcordia and MIL standards, to verify the product's environmental robustness (see Table 8.9).

One important characterization test is the bias/adhesion mirror mapping (BAMM) [41]. It involved the statistical analysis of the number of mirrors that land with increasing applied bias while holding the other operational parameters constant. The result is a distribution curve for mirror landing whose tightness (i.e., spread in voltage) is a measure of process variability that led the engineers to further optimization of the hinge design and voltage drive waveform. The BAMM test also uncovered another weakness: the bias voltage at landing decreased from 16V to about 14V after 2,000 hours of operation. The finding and the data became useful in implementing additional improvements and developing models for lifetime prediction [42].

Table 8.9 A Summary of the Environmental Tests Performed in Assessing the Reliability of the DMD

Environmental Test	*Details*	*Duration*
Storage life (cold/hot)	−55°C to +100°C, no applied power	1,000 hours
Temperature cycling	–55°C to +125°C, air to air, fine/gross leaks	1,000 cycles
Thermal shock	–55°C to +125°C, liquid to liquid	200 cycles
Unbiased humidity	+85°C/85% RH, no applied power	1,000 cycles
Electrostatic discharge	Human body model, 1 positive, 1 negative at 2,000V	
Latch up	25°C, ±300 mA	
Ultraviolet light sensitivity	25°C, ultraviolet exposure	1,000 hours
Sequence 1	1,500G mechanical shock, Y direction only	
	Vibration, 20G from 20 to 2,000 Hz	
	Constant acceleration, 10,000G, Y only	
Sequence 2	Thermal shock, –55°C to +125°C	15 cycles
	Temperature cycling, –55°C to +125°C	100 cycles
	Moisture resistance	10 days

(*Source:* [41].)

Another valuable test and characterization method is the *solution space* technique [41]. In this case, many parameters were controllably varied and plotted in two or more dimensions with the intent of visualizing interrelationships between the variables. An acceptable solution space is one where overall mirror performance is satisfactory under all combinations of operating conditions. The test is performed before and after accelerated aging to gauge the robustness of the solution space and to identify the parameters that were most sensitive to aging. This method also yielded improved hinge designs and electrical drive waveform.

More than two thirds of all failures that affect the DMD micromirror are traced to a particle defect [42], either on the surface of the mirror or underneath it. A particle on the surface affects the rotation dynamics and optical properties of the mirror. A particle below it may prevent mirror movement or cause an electrical short. Particle defects during lithography and etching can damage the hinge. Particle reduction is an important aspect of process control, and, much as in the integrated circuit industry, it greatly impacts yield. The remaining failures are attributed to hinge and mirror mechanics, including hinge fatigue and memory and stiction to the landing electrode.

The hinge is a thin layer (~ 75 nm) of an aluminum alloy (98.8% Al, 1% Si, 0.2% Ti) [43]. To assess its sensitivity to fatigue, Texas Instruments performed accelerated testing by switching the mirrors more rapidly than normal (once every 20 μs). Tests over nearly five million mirrors on nine different DMD dice have accumulated more than 3×10^{12} cycles per mirror without any evidence of fatigue. Naturally, Texas Instruments has been successful in maintaining tight process control over the deposition step and alloy material to result in such consistency in the reliability. Tests, however, demonstrated that hinge memory is a more serious reliability hazard. When a mirror is operated in the same direction for a long period of time, it exhibits a residual tilt in that direction when all bias voltages are removed, due to a permanent deformation in the hinge. This effect is known as hinge memory. When the residual tilt exceeds 3.5° (the full tilt angle of the mirror under operation is 10°), it creates an imbalance in the separation gaps under the mirror, and the electrostatic force on the side with the large gap becomes insufficient to overcome the

permanent twist in the hinge. The pixel appears damaged to the user. Evidence points to metal creep of the hinge material as the source of this effect, with strong dependence on operating temperature and duty cycle. The latter is the percentage of time the mirror lands on one side relative to the other. For example, a 95/5-duty cycle means that the mirror lands on one side 95% of the time, and one the other side the remaining 5%. Tests have shown that duty cycles near or at 50/50 exhibit no hinge memory, but the effect is pronounced at larger duty cycles and is further exacerbated by temperature under worst-case operating conditions (65°C). Duty cycles characteristic of real-life display images tend to be imbalanced (varying between 75/25 and 85/15), thus making hinge memory a limiting factor of lifetime. Early results showed a lifetime of 1,000 hours under worst-case conditions of 65°C and 95/5 duty cycle. A discovery that baking the hinge alloy during fabrication at 150°C for 12 to 16 hours alleviated the tendency to creep by annealing intrinsic stresses and passivating the metal surface [44]. This contributed to a five-fold increase in lifetime. The bake cycle and other additional improvements increased the worst-case lifetime to 10,000 hours, which extrapolates to better than 200,000 hours under normal operating temperatures (<45°C) and duty cycle (75/25 to 85/15) [42].

A last failure mode is the stiction of the yoke to the landing electrode. Stiction remains difficult to predict, but it is believed that contamination of the surface of the landing electrode is the major cause. An innovative solution implemented four spring tips (see Figure 8.19) at the landing corners of the yoke. As the mirror structure tilts and the spring tips come into contact with the landing electrodes, the springs deform and potential energy is stored in them. As soon as the applied bias is removed, the springs push the yoke and the mirror structure off the surface.

Summary

Packaging of MEMS is an art rather than a science. The diversity of MEMS applications places a significant burden on packaging. Standards do not exist in MEMS packaging; rather, the industry has adopted standards and methods from the integrated circuit industry and modified them as necessary. This chapter reviewed the basic considerations of MEMS packaging and introduced three widely accepted packaging approaches: ceramic, metal, and plastic. Basic concepts for reliability are also introduced.

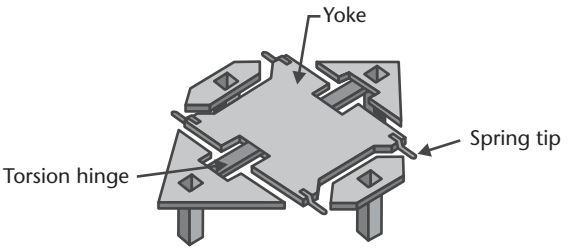


Figure 8.19 An illustration of the middle structure in a DMD showing the spring tips. Their role is to push the mirror off the surface of the landing electrode upon removal of the bias voltage, thus reducing the risk of stiction.

References

[1] Lau, J. H., et al., *Electronic Packaging: Design, Materials, Process, and Reliability*, New York: McGraw-Hill, 1998, pp. 111–193.

[2] Beach, W. F., T. M. Austin, and R. Olson, "Parylene Coatings," in *Electronic Materials Handbook: Volume 1, Packaging*, M. L. Minges, C. A. Dostal, and M. S. Woods (eds.), Materials Park, OH: ASM International, 1989, pp. 789–801.

[3] Monk, D. J., et al., "Media Compatible Packaging and Environmental Testing of Barrier Coating Encapsulated Silicon Pressure Sensors," *Tech. Digest Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, June 3–6, 1996, pp. 36–41.

[4] Flannery, A. F., et al., "PECVD Silicon Carbide for Micromachined Transducers," *Proc. 1997 Int. Conf. on Solid-State Sensors and Actuators*, Vol. 1, Chicago, IL, June 16–19, 1997, pp. 217–220.

[5] Burri, M., "Calibration-Free Pressure Sensor System," Application Note AN1097, Motorola Sensor Device Data book, 1995.

[6] Sergent, J. E., "The Hybrid Microelectronics Technology," in *Electronic Packaging & Interconnection Handbook*, 2nd ed., C. A. Harper (ed.), New York: McGraw-Hill, 1997, pp. 7.10–7.26.

[7] Hwang, J. S., "Solder Technologies for Electronic Packaging," in *Electronic Packaging & Interconnection Handbook*, 2nd ed., C. A. Harper (ed.), New York: McGraw-Hill, 1997, pp. 5.4–5.20.

[8] Striny, K. M., "Assembly Techniques and Packaging of VLSI Devices," in *VLSI Technology*, 2nd ed., S. M. Sze (ed.), New York: McGraw-Hill, 1988, pp. 566-611.

[9] Dow Corning Corporation, Midland, MI 48686-0994, http://www.dowcorning.com/html/industries/electronics/index.html.

[10] *Bonding Handbook and General Catalog*, Kulicke and Soffa Industries, Inc., Willow Grove, PA, 1990.

[11] *Flip Chip Technologies*, J. H. Lau (ed.), New York: McGraw-Hill, 1996.

[12] Jaeggi, D., et al., "Novel Interconnection Technologies for Integrated Microfluidic Systems," *Tech. Digest Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, June 8–11, 1998, pp. 112–115.

[13] VerLee, D., et al., "Fluid Circuit Technology: Integrated Interconnect Technology for Miniature Fluidic Devices," *Tech. Digest Solid-State Sensor and Actuator Workshop*, Hilton Head Island, SC, June 3–6, 1996, pp. 9–14.

[14] Borgeson, P., and P. A. Kondos, "Packaging of Single Mode Laser Diodes," *Proc. of the 2002 ASME International Mechanical Engineering Congress and Exposition*, November 17–22, 2002, New Orleans, LA, EPP-Vol. 2, pp. 471–475.

[15] Harvilchuck, L.A., et al., "Passive Optoelectronics Component Manufacturing," *Proc. of the 2002 ASME International Mechanical Engineering Congress and Exposition*, November 17–22, 2002, New Orleans, LA, EPP-Vol. 2, pp. 465–469.

[16] Ogura, G., "Hermetic Packaging: A Neglected 'Holy Grail'?," *Lightwave Magazine*, October 2001.

[17] Eddy, D. S., and D. R. Sparks, "Application of MEMS Technology in Automotive Sensors and Actuators," in *Integrated Sensors, Microactuators, & Microsystems* (*MEMS*), K. D. Wise (ed.), *Proceedings of the IEEE,* Vol. 86, No. 8, August 1998, pp. 1750–1751.

[18] U.S. Patent 5,545,912, August 13, 1996.

[19] Ginsberg, G. L., and D. P. Schnorr, *Multichip Modules and Related Technologies*, New York: McGraw-Hill, 1994.

[20] *Ball Grid Array Technology*, J. H. Lau (ed.), New York: McGraw-Hill, 1995.

[21] NPC-107 data sheet, GE NovaSensor, 1055 Mission Court, Fremont, CA 94539, http://www.novasensor.com.

[22] Faris, J., and T. Kocian, "DMD™ Packages—Evolution and Strategy," *TI Technical Journal*, July–September 1998, pp. 87–94.

[23] Bonner, J. K., "Surface Mount Technology," in *Electronic Packaging & Interconnection Handbook*, 2nd ed., C. A. Harper (ed.), New York: McGraw-Hill, 1997, pp. 9.50–9.51.

[24] Cohn, C., and M. T. Shih, "Packaging and Interconnection of Integrated Circuits," in *Electronic Packaging & Interconnection Handbook*, 2nd ed., C. A. Harper (ed.), New York: McGraw-Hill, 1997, pp. 6.14–6.17.

[25] Cohn, C., and M. T. Shih, "Packaging and Interconnection of Integrated Circuits," in *Electronic Packaging & Interconnection Handbook*, 2nd ed., C. A. Harper (ed.), New York: McGraw-Hill, 1997, pp. 6.46–6.56.

[26] Available from The U.S. Department of Defense, Single Stock Point for Military Specifications, Standards and Related Publications, Bldg. 4/D, 700 Robbins Avenue, Philadelphia, PA, 19111–5094, http://www.dodssp.daps.mil.

[27] Telcordia® Technologies, One Telcordia Drive, Piscataway, NJ 08854, http://www.telcordia.com.

[28] Bājenescu, T. I., and M. I. Bâzu, *Reliability of Electronic Components*, New York: Springer, 1999, pp. 9–20.

[29] Kececioglu, D., *Reliability Engineering Handbook*, Vol. 1, Engelwood Cliffs, NJ: Prentice Hall, 1991, pp. 107–323.

[30] Glasstone, S., K. J. Laidler, and H. Eyring, *The Theory of Rate Processes*, New York: McGraw-Hill, 1941.

[31] DeGroot, M. H., and M. J. Schervish, *Probability and Statistics*, 3rd ed., Boston, MA: Pearson Addison Wesley, 2002.

[32] *NIST/SEMATECH e-Handbook of Statistical Methods*, Section 8.1.5.2, 2003, http://www.itl.nist.gov/div898/handbook.

[33] Nelson, W., *Accelerated Testing: Statistical Models, Test Plans, and Data Analyses*, New York: Wiley, 1990, pp. 51–106, 521–543.

[34] *Accelerated Life Testing Reference*, ReliaSoft Corporation, 115 South Sherwood Village Dr., Tucson, AZ 85710, 2001, http://www.weibull.com.

[35] Allameh, S. M., et al., "Surface Topography Evolution and Fatigue Fracture in Polysilicon MEMS Structures," *Journal of Microelectromechanical Systems*, Vol. 12, No. 3, June 2003, pp. 313–324.

[36] Tanner, D. M., et al., "MEMS Reliability in Shock Environments," in *Proceedings IEEE International Reliability Physics Symposium*, San Jose, CA, April 2000, pp. 129–138.

[37] Reynolds, J. K., et al., "Packaging a Piezoresistive Pressure Sensor to Measure Low Absolute Pressures over a Wide Sub-Zero Temperature Range," *Sensors and Actuators A: Physical*, Vol. 83, 2000, pp. 142–149.

[38] Srikar, V. T., and S. D. Senturia, "The Reliability of Microelectromechanical Systems (MEMS) in Shock Environments," *Journal of Microelectromechanical Systems*, Vol. 11, No. 3, June 2002, pp. 206–214.

[39] Hyman, D., and M. Mehregany, "Contact Physics of Gold Microcontacts for MEMS Switches," *IEEE Transactions on Components and Packaging Technology*, Vol. 22, September, 1999, pp. 357–364.

[40] Goldsmith, C., et al., "RF MEMS: Benefits and Challenges of an Evolving Switch Technology," in the 23rd Annual Technical Digest, *Gallium Arsenide Integrated Circuit (GaAs IC) Symposium*, October 21–24 , 2001, pp. 147–148.

[41] Douglass, M. R., "DMD Reliability: A MEMS Success Story," in *Reliability, Testing and Characterization of MEMS/MOEMS II,* R. Ramesham and D. M. Tanner (eds.), *Proceeding of the SPIE,* Vol. 4980, January 2003, pp. 1–11.

[42] Douglass, M. R., "Lifetime Estimates and Unique Failure Mechanisms of the Digital Micromirror Device (DMD)," *Proceedings of IEEE International Reliability Physics Symposium*, 1998, pp. 9–16.

[43] U.S. Patents 4,662,746, May 5, 1987, and 5,083,857, January 28, 1992.

[44] U.S. Patent 6,204,085, March 20, 2001.

Selected Bibliography

Electronic Packaging & Interconnection Handbook, 2nd ed., C. A. Harper (ed.), New York: McGraw-Hill, 1997.

Harper, C. A., and M. B. Miller, *Electronic Packaging, Microelectronics and Interconnection Dictionary*, New York: McGraw Hill, 1993.

Kececioglu, D., *Reliability Engineering Handbook*, Vols. 1 and 2, Englewood Cliffs, NJ: Prentice Hall, 1991.

Lau, J. H., et al., *Electronic Packaging: Design, Materials, Process, and Reliability*, New York: McGraw-Hill, 1998.

Manzione, L. T., *Plastic Packaging of Microelectronic Devices*, New York: Van Nostrand Reinhold, 1990.

Striny, K. M., "Assembly Techniques and Packaging of VLSI Devices," in *VLSI Technology*, 2nd ed., S. M. Sze (ed.), New York: McGraw-Hill, 1988, pp. 566–611.

Glossary

Accelerated life testing Also known as accelerated ageing, it is a methodology in reliability that subjects a product to a high degree of operational and environmental stresses, such as high temperature or voltage, in order to precipitate failure modes that would normally manifest themselves in later stages of the product's life.

Action potential A temporary change in the electrical voltage across the cellular membrane of a nerve or muscle cell. Action potentials occur when the cell is stimulated, especially by a nerve impulse. They form the mechanism by which sensory and motor functions are transmitted across the neural system.

Amorphous silicon Silicon lacking a preferred crystalline orientation, typically consisting of extremely fine grains each measuring a few nanometers in size.

Amplification In biochemistry, it is the process of making a large number of identical copies of a DNA fragment. In electronics, it is the process of increasing the magnitude of an electrical voltage.

Anodic bonding A process to bond silicon to glass, specifically Pyrex™ 7740 or equivalent.

Application-specific integrated circuit (ASIC) An electronic integrated circuit with a functionality designed specifically for one particular application (e.g., the detection of minute changes in capacitance and conversion into an output voltage).

Bandpass *see* filter.

Bandwidth The extent of the frequency response of a linear system. It is numerically defined as the difference between two corner frequencies where the system gain is 3 dB below the maximal gain. An input signal with a frequency content below or above the corner frequencies is severely attenuated by the system.

Bimetallic actuation The resulting motion when a stack of two materials having dissimilar coefficients of thermal expansion is heated. One material expands more than the other, giving rise to bending stresses. The amount of bending is proportional to the temperature of the stack and the difference in coefficients of thermal expansion.

BGA An acronym for ball grid array. A type of advanced ceramic package for integrated circuits consisting of arrays of solder balls instead of electrical pins.

Bond pad A metal area on a die or wafer to which a gold or aluminum wire is bonded. The wire and bond pad provide electrical connectivity between electrical components on the die and electronic circuitry external to the die.

Bragg grating A periodic variation in the index of refraction within the core of an optical fiber or a waveguide. It reflects light at its resonant wavelength that is proportional to the periodic spacing of the grating.

Brownian noise *see* noise.

Bulk micromachining A class of micromachining processes that remove material from a substrate, yielding micromechanical structures with thicknesses in the tens or hundreds of micrometers. It usually refers to the orientation-dependent etching or deep reactive ion etching of silicon.

Burn-in A methodology in reliability testing that involves operating a device under extreme operating and environmental conditions in order to accelerate the life of the product and precipitate an early failure.

Butterfly package A metal package that consists of a tub-like housing with electrical leads on two sides. A butterfly package may also include a feed-through tube to allow for a fiber optical connection.

CERDIP An acronym for ceramic dual-in-line package for integrated circuits. It consists of a rectangular pressed ceramic body with pins on two opposites. A ceramic cap is glass sealed to the body.

CERQUAD An acronym for ceramic quad-flat pack. Similar to the CERDIP but has pins on all sides.

Characteristic impedance The input impedance of a transmission line. Intuitively, it is the impedance of the load that, when attached to the terminal ends of a transmission line, makes the line appear infinitely long. Characteristic impedance is of great importance for good low-loss transmission. If the characteristic impedance of a transmission line is mismatched to the load at the output terminals, then some of the signal is reflected back at the junction causing loss and other undesirable effects. It is a complex number with resistive and reactive components. It is independent of the length of the transmission line but depends on frequency.

Check valve A valve that permits fluid flow only in one direction.

Chemical vapor deposition (CVD) A process based on the principle of initiating a chemical reaction in a vacuum chamber, resulting in the deposition of a reacted species on a heated substrate. Materials that can be deposited by CVD include polysilicon, silicon oxide, and silicon nitride.

Chromaticity A property of light that relates to the number of wavelengths (colors) available within an optical beam. A laser is a highly chromatic source because it contains essentially one main wavelength.

CMOS An acronym for complementary metal oxide semiconductor. A class of electronic devices made of silicon and associated fabrication processes common in integrated circuits.

Coefficient of thermal expansion (CTE) The rate of change in length of an object as a function of temperature. In general, $CTE = (\Delta L/L)/\Delta T$ where $(\Delta L/L)$ is the fractional change in length corresponding to a ΔT change in temperature. It is measured in inverse units of temperature (/°C or ppm/K).

Coherence A property of light that describes the statistical correlation between the phases of optical wavefronts (i.e., the degree to which the waves of the radiation are vibrating in phase). Temporal coherence relates to the phase correlation between different instants in time and is a necessary characteristic to obtain interference. Spatial coherence relates to the phase correlation between different points in space.

Complementarity In biochemistry, it is the specific affinity for binding between the purines (adenine and guanine) with pyrimidines (thymine and cytosine). *See* nucleotide.

Condenser microphone see capacitor microphone.

Coriolis effect Physical effect responsible for the deflection of objects moving on the surface of a rotating body such as Earth.

Corner frequency *see* filter**.**

Decibel (dB) A unit to measure the relative difference in intensity of a physical, electrical, optical, or acoustical signal (e.g., voltage, electric field, or pressure). It is defined as 20 times the 10-base logarithm of the *intensity* ratio, or, because power is proportional to the square of the intensity, 10 times the 10-base logarithm of the *power* ratio. For example, 40 dB is equivalent to a ratio of 100 between the highest and lowest intensities in the range. The same ratio in power would, however, be equivalent to 20 dB.

Degeneracy *see* frequency degeneracy.

Detector Also photodetector, it is a semiconductor sensor that converts the optical intensity in a light beam into an electrical current. A quadrant detector is made of four adjoining detectors and detects the spatial position of a light beam.

Die Also chip, a common term in microfabrication technology indicating a small piece of semiconductor or glass cut or diced from a much larger wafer.

Diffusion In semiconductor fabrication, it is the process to controllably spread or diffuse impurity dopant atoms in silicon or a semiconductor. Diffused resistors are resistors made of one type of doping (e.g., *p*-type) and embedded in silicon having a background doping of the opposite type (e.g., *n*-type).

Diffraction In optics, it is the effect of an incident light beam separating into multiple beams, each called a higher diffraction order upon striking an optical element with features that approach in size the wavelength of the incident beam.

DIP An acronym for dual-in-line type of package. Made of ceramic or plastic, it is rectangular in shape with pins (leads) on its two long sides.

Dipole Also electric dipole, it is the electric field created by two charges of equal magnitude but opposite polarity and separated by a small distance.

Distributed-feedback laser (DFB) A category of integrated semiconductor lasers that utilize a Bragg grating to stabilize their output optical frequency. They are widely used in optical fiber telecommunications.

Doping Also known as impurity doping. A process of introducing into a semiconductor material impurities or foreign atoms—dopants—in relatively dilute concentrations ($10^{13} \sim 10^{20}$ cm^{-3}). The impurities alter the electrical properties of the

semiconductor by adding electrons or holes (carriers). At or near room temperature, the carrier concentration is largely equal to the dopant concentration. If the doping is *n*-type, then there is an excess of electrons. Conversely, a *p*-type material has an excess of holes. The material is electrically more conductive at higher doping levels. A *p*-type region in direct contact with an *n*-type forms a *p*-*n* diode, which passes current only in one direction: from the *p*-type to the *n*-type. Arsenic and phosphorous are common *n*-type dopants in silicon; boron is a *p*-type dopant in silicon.

Duality A general abstract concept that pairs equivalent parameters from distinct physical systems on the basis of energy arguments. Duality is frequently invoked between mechanical, thermal, and electrical systems. For example, a spring in a mechanical system is dual to a capacitor because they both store potential energy. In general, duality pairs mass to inductance, spring constant to the inverse of capacitance, coefficient of viscous damping to resistance, mechanical displacement to charge, velocity to electrical current, and applied force to applied voltage.

EEPROM An acronym for electrically erasable and programmable read only memory, a type of read-only electronic memory that can be erased and reprogrammed using high-voltage electrical pulses.

Electromagnetic interference (EMI) Undesirable interference with electronic signals of an electromagnetic nature. Sources of EMI include solar eruptions, radio signals, and nuclear explosions.

Electrophoresis In chemistry, it is the migration of charged or polar molecules in suspension in a solution under the effect of an externally applied electric field. It is useful for the separation of dissimilar molecules and analysis of their molecular structure based on their rate of movement.

Epitaxy Chemical process to grow a thin crystalline layer on top of a crystalline substrate. The grown layer generally has the same crystalline orientation as the substrate.

Etalon Also known as a Fabry-Perot cavity or etalon, it is an optical resonator bounded by two partially reflective surfaces. At resonance, the length of the cavity equals an integer number of half wavelengths. The etalon is a useful optical filtering element.

Eutectic point At their eutectic point, alloys have identical solidus and liquidus temperatures. The melting temperature of a eutectic alloy is lower than that of any other alloy composed of the same constituents in different proportions. *See* liquidus temperature.

Exponential probability density function A commonly used probability distribution in reliability applicable to components with a constant failure rate.

Filter An electrical or optical circuit that selectively blocks the transmission of certain frequencies (or wavelengths). The transition frequencies defining the bands of transmission are known as corner frequencies. A lowpass filter blocks high frequencies but permits the transmission of low frequencies. A highpass filter performs the opposite function of a lowpass filter. A bandpass filter allows the transmission of

frequencies in a mid-band range but blocks the transmission of frequencies above or below this band—outside of the corner frequencies.

Finesse A measure of optical loss in an optical resonator, such as an etalon. Imperfections in the reflective surfaces that bound the resonator and in the material between them degrade the finesse. It is the ratio of the free spectral range (FSR), the spectral separation between two adjacent transmission peaks of the resonator, to the full width at half maximum (FWHM), the spectral width of the transmission peak at its 50% point.

FIT An acronym for failure in time, it is a measure of reliability equal to 10^{-9} failure/hr.

Foundry A service facility capable of prototyping and fabricating semiconductor circuits or MEMS. Foundry services typically offer a set of standard fabrication processes. A few provide custom design services.

Frequency degeneracy The situation when two or more resonant modes oscillate at exactly the same frequency. When a number of identical oscillators are coupled with each other, their frequencies become degenerate. The coupling generally lifts this degeneracy by splitting the frequencies apart. The amount of separation depends on the strength of the coupling.

Genomics The study of genes and their functions.

Grating Also known as diffraction grating, it consists of a repeating array of lines spaced apart by a distance close to that of the wavelength of light. Incoming light is diffracted by the grating into a multitude of higher orders that form an angle with the incident light beam. A grating can be either transmissive or reflective. *See* Bragg grating.

Helmholtz cavity Also known as a Helmholtz resonator, in acoustics it is a hollow air-filled cavity having an inlet opening for sound and an outlet. The cavity is an acoustic oscillator with a characteristic resonant frequency determined by both the air volume and the geometry of the inlet and outlet ports. It is commonly used in acoustics for frequency tuning.

Hole In physics, it is a vacant position in a semiconductor left by the absence of an electron. The concept is analogous to a bubble in water left by the absence of liquid. A hole is a carrier of positive electric charge and participates in electric conduction.

Hybridization In biochemistry, it is the process when two DNA strands having complementary sequences of nucleotides match up and bind with each other.

Impedance A measure of the total resistance to electrical current flow. In acoustics, it is a measure of the total resistance to the propagation of acoustic pressure waves through a medium. Impedance is generally a complex number.

Insertion loss In a linear system, such as a filter, it is the attenuation, measured in decibels, of an input signal with frequency content within the system bandwidth. Ideally, it is zero.

Ion implantation A high-energy process capable of embedding impurity dopant atoms within the surface of a semiconductor substrate. It is usually followed by a

high-temperature diffusion or anneal step. Implantation is useful in the doping of piezoresistors, embedded electrical interconnects in a silicon substrate, and thin polysilicon films.

International Telecommunication Union (ITU) The ITU, headquartered in Geneva, Switzerland, is an international organization within the United Nations, where governments and companies coordinate global telecommunication networks, services, and standards.

Liquidus temperature In metallurgy, the phase state of an alloy changes with temperature, pressure, and mole percentage of its constituents. At a constant pressure, there are three distinct regions in the phase diagram. At low temperatures, the alloy is a solid. At high temperatures, it is a liquid. A third intermediate region defines a plastic-like, mixed liquid and solid state. The dividing line between the liquid state and the plastic-like state is the liquidus line. The dividing line between the solid state and the plastic-like state is the solidus line. For each molar composition, there is a liquidus and a solidus temperature. At the eutectic composition, the two temperatures are identical; in other words, the two lines coalesce and the plastic-like phase vanishes.

Lithography A process common in microfabrication for delineating a pattern image in a photosensitive polymer. The polymer, or photoresist, can then be used as a masking layer to transfer the pattern into the underlying substrate.

Lorentz force In physics, it is the force on a current-conducting element inside a magnetic field. Its magnitude is equal to the current multiplied by the magnetic flux density and the length of the conductor. Its direction is orthogonal to both the magnetic field and electric current. For a free charge, the force is equal to the product of the charge, the velocity, and the magentic flux density.

LPCVD An acronym for low-pressure chemical-vapor deposition.

MESFET An acronym for metal-semiconductor field effect transistor. A type of electronic transistor useful for operation at very high frequencies. It is very common in electronic circuits made of gallium arsenide (GaAs).

Microelectromechanical systems (MEMS) A generic descriptive term, common in the United States, for a broad technology with the objective of miniaturizing complex systems by integrating a diverse set of functions into a small package or a single die.

Micromachining A term describing the set of design and fabrication tools for the machining of microstructures and very small mechanical features in a substrate frequently made of silicon.

Microwave The electromagnetic frequency range variously defined as starting at about 1 GHz and extending to 30 GHz or even 3 THz (the far infrared). The range of 30 GHz to 300 GHz, with wavelengths in vacuum of 1 cm to 1 mm, is referred to as the millimeter-wave range.

MTTF An acronym for mean time to failures for nonrepairable systems. A typical figure for robust electronic equipment is in excess of 500,000 hours. It means

that of all equipment tested, one failure occurred after 500,000 hours of testing. MTTF is usually predicted using statistical methods.

MTBF An acronym for mean time between failures. It is used to measure the mean time between successive failures, assuming the system can be repaired to the same state prior to the failure.

Multi-chip modules (MCM) A type of high-density packaging approach common in the integrated circuit industry that involves electrically connecting a number of dice on the same substrate.

***n*-type doping** *see* doping.

Noise A random disturbance in an electrical, optical, acoustic, or mechanical signal. It frequently determines the resolution of a sensor. It is characterized by its spectral density, which represents the magnitude in an infinitesimally small frequency band. The power in the noise signal is proportional to frequency bandwidth; hence, the voltage magnitude is proportional to the square root of bandwidth. Noise originates from a multitude of sources. The most common is electrical interference noise and may be filtered. Thermal—or Brownian—and *1/f* noise are fundamental physical entities. Thermal noise originates from physical mechanisms where energy is converted to heat (e.g., electrical resistance or mechanical friction). It is white in nature, meaning its spectral energy density is constant over frequency. Thermal noise increases with temperature. In contrast, *1/f* noise, as the name implies, has a spectral density that decreases at higher frequencies. It is common in electronic circuits, and originates from crystal imperfections that momentarily trap electrons (hence the frequency dependence). The corner frequency in a noise spectrum is the frequency where *1/f* and thermal noise are equal. A common measure of electrical noise is the magnitude of the spectral density at a particular frequency, given in $V/\sqrt{Hz}$.

Nucleotide The building block of deoxyribonucleic acid (DNA). It can be any of adenine (A), cytosine (C), guanine (G), or thymine (T). The sequence of nucleotides in DNA defines the genetic code. An oligonucleotide is a molecule composed of 25 or fewer nucleotides.

Numerical aperture (NA) In optics, it is the sine of the angle of the largest cone of meridional rays (rays that cross the optic axis) that can enter or leave an optical system or element multiplied by the index of refraction. For NA$<$0.25, it is approximately equal to one-half the inverse of the *f*-number and a measure of the aperture size. The numerical aperture is always less than one in air.

Oxidation Chemical process by which the atoms of an element lose electrons. In an aqueous solution, neutral atoms become positive ions.

Optoelectronic Family of devices and components that merge both optical and electronic functions. They include lasers, optical detectors, modulators, optical amplifiers, and integrated assemblies merging all such building blocks.

PECVD An acronym of plasma-enhanced chemical-vapor deposition.

Phase quadrature *see* quadrature.

Piezoelectricity The property exhibited by a class of materials to develop a voltage in response to applied mechanical stress or pressure. Conversely, an externally applied electrical voltage strains and deforms the material.

Piezoresistivity The property of a certain class of materials, including impurity doped silicon, to change their electrical resistivity in response to mechanical stress.

Plasma etching A class of etch processes capable of selectively removing material, including silicon, by chemical reaction with one or more gases. The reactive gases are ionized in a plasma inside a vacuum chamber by means of electrical or electromagnetic energy. A plasma is an electrically neutral, highly ionized gas composed of ions, electrons, and neutral particles.

Polysilicon Abbreviation for polycrystalline silicon. An aggregate of small crystalline grains of silicon, each with a different preferred orientation. The grains may vary in dimensions from a few nanometers to a few micrometers.

Polymerase chain reaction (PCR) In biochemistry, it is an amplification process invented in the 1980s for creating billions of identical replicas of a DNA fragment.

Primitive unit The smallest repeating block of a crystal lattice.

Proteomics The study of the full set of proteins encoded by a genome.

***p*-type doping** *see* doping.

***p-n* diode** *see* doping.

Quadrature The situation when two periodic signals of same frequency, f_L, are out of phase by a quarter of a cycle, or 90°. For example, sine and cosine waveforms are in quadrature (or phase quadrature). One important application is in communications and RF circuits. Separation is possible by heterodyning (multiplication) with another signal of frequency f_R. The amplitudes at the two new frequencies, $(f_R + f_L)$ and $(f_R - f_L)$, are proportional to the amplitude sum and difference, respectively, of the signals in quadrature.

Quality factor Physically, it is 2π multiplied by the ratio of the energy stored to the energy lost during one oscillation cycle of a resonator. It arises from energy loss mechanisms, such as resistive heating in electrical networks, or viscous damping or friction at grain boundaries in mechanical systems. Practically, it is the ratio of the resonant frequency to the spectral bandwidth measured at 3 dB below the resonant peak of an electrical or mechanical system. The sharper the resonance is, the higher the quality factor gets. It is a measure of the frequency stability of oscillators. In an RLC electrical circuit, it is equal to $\sqrt{LC}/R$.

Radio Frequency (RF) Also known as radio spectrum, it is a general term that refers to the portion of the electromagnetic spectrum between 3 kHz and 300 GHz. In the United States, the Department of Commerce and the Federal Communications Commission are responsible for regulating the frequency allocations in this spectrum.

Reduction Chemical process by which the atoms of an element gain electrons and increases their negative valence. Reduction neutralizes positive ions in an aqueous solution.

Resistivity The material property, usually measured in Ω•cm, that, along with the physical dimensions, determines the resistance of a resistor in Ω.

Sacrificial etching A micromachining processing method in which an intermediate layer sandwiched between two layers of a different material is preferentially (sacrificially) etched and selectively removed. Usually, the etch selectivity is high between the intermediate layer and the two sandwich layers. The purpose of the sacrificial layer is to mechanically release one or both of the sandwich layers. Silicon oxide is a commonly used sacrificial layer.

Silicon fusion bonding Also known as silicon direct bonding, it is a process to fuse or bond together two silicon substrates. The bond is strong, generally occurring at the molecular level.

Silicon on insulator (SOI) Substrates consisting of a thin layer of silicon dioxide, typically 0.5 to 2 μm thick, sandwiched between two crystalline silicon layers. The silicon dioxide is known as buried oxide. One method to fabricate SOI substrates is by silicon fusion bonding a silicon wafer with a thin layer of silicon dioxide on its surface to a bare silicon wafer, then thinning one of the silicon substrates. SOI is a well-proven technology for the fabrication of CMOS electronic circuits suitable for high-temperature operation (up to 300°C), as well as for high-voltage (>100V) and high-frequency (>10 GHz) applications.

Sheet resistance The resistance of one square of material in units of Ω per square (Ω/□). It is equal to resistivity divided by the thickness of the material. For thick-film resistors, it is generally implicit that the unit thickness is one mil (25.4 μm).

Solidus temperature *see* liquidus temperature.

Sound Power Level (SPL) Sound pressure, in decibels, measured in reference to a base sound pressure of 20 μPa in air. The reference is usually 1 Pa in water.

Sputtering A process to deposit a thin film on the surface of a substrate. It involves the removal of material from a target by ion bombardment and subsequent redeposition on the substrate.

SRAM An acronym for static random access memory, a type of electronic memory that can be arbitrarily addressed. Unlike EEPROM, it cannot hold the data once electrical power is turned off.

Stiction A well-known problem resulting from the geometrical scaling law when surface adhesion forces are larger than the mechanical restoring forces of a suspended micromachined structure.

Strain In mechanics, a deformation produced by stress. In a beam, it is equal to the change in length divided by the original beam length.

Surface micromachining A class of fabrication processes yielding micromechanical structures which are only a few micrometers thick, usually involving the removal of a sacrificial layer.

Surface mount technology (SMT) An advanced electronic packaging technology in which the type of packages are particularly small so that they can be soldered in high density on the surface of a printed-circuit board.

SVGA *see* VGA.

SXGA *see* VGA.

TEC *see* Thermoelectric cooler.

Temperature coefficient of expansion (TCE) *see* coefficient of thermal expansion.

Temperature coefficient of resistance (TCR) The rate of increase in resistance as a function of temperature. In general, $TCR = (\Delta R/R)/\Delta T$, where $(\Delta R/R)$ is the fractional change in resistance corresponding to a ΔT change in temperature. It is measured in inverse units of temperature (/°C).

Thermocompression bond A bonding process involving the melting of an intermediate layer between two substrates pressed against each other. Frequently, the intermediate layer is made of glass or gold.

Thermoelectric cooler (TEC) Also known as a Peltier cooler, it is a device that is capable of transferring heat from one side to another. It is common in controlling the temperature of small dice or objects. The device consists of many semiconductor junctions in parallel that use the Peltier effect.

VGA An acronym for video graphics adapter, it identifies displays with a resolution of 640 × 480 pixels. SVGA, XGA, and SXGA denote displays with resolutions of 800 × 600, 1,024 × 768, and 1,280 × 1,024 pixels, respectively.

Wavelength A characteristic of a wave, it represents the distance between identical points in two adjacent cycles of an oscillating waveform. In electromagnetic waves, wavelength and frequency are inversely related by the velocity of the wave. In vacuum, the velocity is the speed of light equal to 3×10^8 m/s. In a transmission line, the velocity depends on the dielectric constant, frequency, and geometry and is in the range of 2×10^8 m/s. Thus, a 2-GHz signal has a wavelength of around 100 mm.

Wavelength division multiplexing (WDM) A protocol in fiber-optic communication in which digital data is multiplexed on different wavelengths in a single fiber. This effectively increases the bandwidth available in one fiber by increasing the number of channels.

Wavelength locker A device used in fiber optical telecommunications that locks the wavelength of a laser to a specific value that is often preset by the ITU.

Wet anisotropic etching Process of etching or removal of material from a silicon substrate with the etch front delineated by crystallographic planes. Potassium hydroxide and tetramethyl ammonium hydroxides are two examples.

Wheatstone bridge An electrical circuit consisting of four resistors forming two branches electrically connected in parallel, with each branch consisting of two resistors electrically in series. It is useful to measure an imbalance in the values of the four resistors.

Wireless RF communication without the use of wires or other transmission lines.

Young's modulus Also known as modulus of elasticity, it is a material constant (in units of pressure) relating mechanical stress to elastic strain. It is indicative of the stiffness of the material. For example, diamond has a very high Young's modulus, whereas soft polymers have low values. It often depends on orientation in crystalline materials.

About the Authors

Nadim Maluf received a B.E. from the American University of Beirut, Lebanon, an M.S. from the California Institute of Technology, and a Ph.D. from Stanford University, all in electrical engineering.

He is currently vice president of marketing at New Focus, a division of Bookham Technology, in San Jose, California. Prior to his current position, he held senior management positions at New Focus and was the head of the R&D department at NovaSensor in Fremont, California. He is also a consulting professor of electrical engineering at Stanford University. He has more than 20 years of industry experience in integrated circuit technology, microelectromechanical systems, sensors and actuators, and their use in medical, automotive, industrial, and telecommunications applications.

Kirt Williams is a senior scientist at Science Applications International Corporations. He has conducted research and development on a variety of micromachined devices, including all-silicon plate valves for high pressure and a wide temperature range, high-Q variable capacitors, optical switches, and thermionic-emission-based ionization pressure sensors and magnetic-field sensors. He is also a short course lecturer on MEMS for the University of California Extension. He received a B.S. with a double major of electrical engineering and materials science and an M.S. and a Ph.D. in electrical engineering, all from the University of California at Berkeley. He is a state-licensed professional engineer in electrical engineering. Dr. Williams has authored or coauthored a number of papers, patents, and a book chapter on microelectromechanical devices and fabrication processes.

Index

T

Recent Titles in the Artech House Microelectromechanical Systems (MEMS) Series

Fundamentals and Applications of Microfluidics, Nam-Trung Nguyen and Steven T. Wereley

Introduction to Microelectromechanical (MEM) Microwave Systems, Héctor J. De Los Santos

An Introduction to Microelectromechanical Systems Engineering, Second Edition, Nadim Maluf and Kirt Williams

MEMS Mechanical Sensors, Stephen Beeby et al.

RF MEMS Circuit Design for Wireless Communications, Héctor J. De Los Santos

For further information on these and other Artech House titles, including previously considered out-of-print books now available through our In-Print-Forever® (IPF®) program, contact:

Artech House
685 Canton Street
Norwood, MA 02062
Phone: 781-769-9750
Fax: 781-769-6334
e-mail: artech@artechhouse.com

Artech House
46 Gillingham Street
London SW1V 1AH UK
Phone: +44 (0)20 7596-8750
Fax: +44 (0)20 7630-0166
e-mail: artech-uk@artechhouse.com

Find us on the World Wide Web at:
www.artechhouse.com