

PETER VAN ZANT

Microchip Fabrication

THIRD EDITION

A PRACTICAL GUIDE TO
SEMICONDUCTOR PROCESSING

Microchip Fabrication

A Practical Guide to Semiconductor Processing

Peter Van Zant

Third Edition

McGraw-Hill

New York San Francisco Washington, D.C. Auckland Bogotá
Caracas Lisbon London Madrid Mexico City Milan
Montreal New Delhi San Juan Singapore
Sydney Tokyo Toronto

Library of Congress Cataloging-in-Publication Data

Van Zant, Peter.

Microchip fabrication : a practical guide to semiconductor processing / Peter Van Zant.—3rd ed.

p. cm.

Includes index.

ISBN 0-07-067250-4 (hardcover)

1. Semiconductors—Design and construction. I. Title.

IN PROCESS

621.3815'2—dc20

96-15729

CIP

McGraw-Hill

A Division of The McGraw-Hill Companies



Copyright © 1997 by The McGraw-Hill Companies, Inc. All rights reserved. Printed 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 data base or retrieval system, without the prior written permission of the publisher.

3 4 5 6 7 8 9 0 DOC/DOC 9 0 1 0 9 8 7

ISBN 0-07-067250-4

The sponsoring editor for this book was Steve Chapman, the editing supervisor was Fred Bernardi, and the production supervisor was Pamela Pelton. It was set in Century Schoolbook by Dina E. John of McGraw-Hill's Professional Book Group composition unit.

Printed and bound by R. R. Donnelley & Sons Company.



This book is printed on recycled, acid-free paper containing a minimum of 50% recycled, de-inked fiber.

Information contained in this work has been obtained by The McGraw-Hill Companies, Inc. ("McGraw-Hill") from sources believed to be reliable. However, neither McGraw-Hill nor its authors guarantees the accuracy or completeness of any information published herein and neither McGraw-Hill nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw-Hill and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

INTERNATIONAL EDITION

Copyright ©1997. Exclusive rights by The McGraw-Hill Companies, Inc. for manufacture and export. This book cannot be re-exported from the country to which it is consigned by McGraw-Hill. The International Edition is not available in North America.

When ordering this title, use ISBN 0-07-114837-X.

Preface to the Third Edition

The first edition of *Microchip Fabrication* came off press in 1984. Feature sizes were still in the micron range, a circuit could be fabricated with a few dozen major steps, and a wafer-fabrication facility could be built for about \$10 million. We knew that Moore's law was racing along like some out-of-control train. Since 1984 it has crashed through the micron feature size barrier, the multimillion transistor chip, and the gigabuck (\$1 billion) wafer-fabrication facility. These pressures have driven wafer fabrication into a dizzying pace of changes.

The pressures tormenting the industry also show up in this edition. It has been our goal to write *Microchip Fabrication* in a way such that fabrication workers or suppliers to wafer-fabrication operations would have a clear, concise first text with which to launch their careers. Exploding new processes and technologies, each more complicated and demanding, have placed new challenges on achieving our goals. For any given process, there are numbers of options and endless variations on those options. *Microchip Fabrication* covers the basics of processing and reflects current manufacturing practices. It is our intention to give the reader a thorough grounding in the problems and current solutions, thus allowing self-education as the industry continues its relentless march to the outer fringes of technology.

As in the second edition, this third edition is organized around the principal processing technologies. Where there are shared processes (such as tube furnaces in diffusion and CVD), the process is described once and referenced in the other sections. This approach has created a lot of forward and backward references. We hope the confusion is tolerable.

We would like to thank the following industry folks who graciously shared their information and experience for this update: Anne Miller

(Semiconductor Services), Don Keenan (Siltec Corporation), Roger Brennan (Solecon), Jerry Scheely (Technical Instruments Company), Mike Wright (Wright Associates), Gene Sullivan (Dryden Engineering), Jim Reynolds (Reynolds Consulting), Don Levinthal (Semiconductor International), Adam Japko (Solid State Technology), Bill Moffet and Jim Reynolds (Yield Engineering Systems), and John Price (In-Spec).

Also, thanks and gratitude to my ever-supportive wife, Mary DeWitt.

Peter Van Zant

Over

In th
and
and
will h

Obje

Upon

1. De
ci
2. W
ar
3. Li
4. Ex
ci
5. Li

Birth

Signi
soluti

rent flow. Capacitors are two-element devices that store charge in a circuit. In some circuits the technology is used to create fuses. Refer to Chap. 14 for an explanation of these concepts and an explanation of how these devices work.

Discrete devices

The announcement in 1947 of the first solid-state transistor launched the solid-state electronics age and the microelectronics industry. The industry is divided into two major segments, the semiconductor industry and the electronics industry. The electronics industry designs and produces a vast number of semiconductor-device-based products, from consumer electronics to computers. It also includes the manufacturers of printed circuit boards.

The semiconductor segment is composed of two major subsegments. One comprises the firms that actually make semiconductor solid state-devices and circuits, using a manufacturing process called *wafer fabrication*. The other subsegment is the cadre of companies that develop and supply the special equipment, materials, and services required for wafer fabrication.

The semiconductor industry was in full swing by the early 1950s supplying devices for transistor radios and transistor-based computers. In this early stage the semiconductor industry produced separate transistors, diodes, resistors, and capacitors (Fig. 1.5), known as *discrete devices*, because they contained only one device per chip. Most discrete devices have less demanding operational and fabrication requirements than integrated circuits. In general, discrete devices are not considered leading edge products. Yet they are required in most sophisticated circuits, and traditionally have accounted for 20 to 30 percent of the dollar volume of all semiconductor devices sold.

Integrated circuits (ICs)

The dominance of discrete devices in solid-state circuits came to an end in 1959. In that year, Jack Kilby, a new engineer at Texas Instruments in Dallas, formed a complete circuit on a single piece of

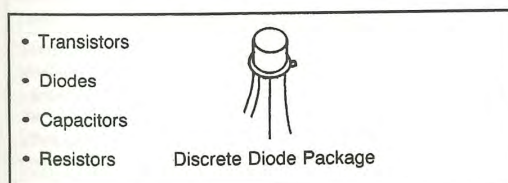


Figure 1.5 Solid-state discrete devices.

the semiconducting material germanium. His invention combined several transistors, diodes, and capacitors (five components total), and used the natural resistance of the germanium chip (called "a bar" by Texas Instruments) as a circuit resistor. This invention was the *integrated circuit*, the first successful integration of a complete circuit in and on the same piece of a semiconducting substrate. Figure 1.6 is a drawing of the Kilby circuit. Note that the devices are connected with individual wires.

The Kilby circuit did not have the form that is prevalent today. Robert Noyce, then at Fairchild Camera, created the final piece of the puzzle. Earlier, Jean Horni, also at Fairchild Camera, had combined a diffused junction transistor on a silicon chip to produce a solid-state transistor with a flat profile (Fig. 1.7). The flattened profile was the outcome of taking advantage of the easily formed natural oxide of silicon, which also happened to be a dielectric (electrical insulator). Horni's transistor used a layer of evaporated aluminum that was patterned into the proper shape to serve as wiring for the device. This technique is called *planar technology*.

Noyce applied this technique to "wire" individual devices together on the silicon wafer surface. The Noyce integrated circuit became *the* model for all integrated circuits. The techniques used not only met

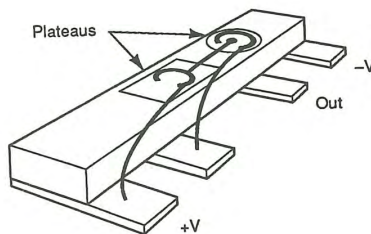


Figure 1.6 Kilby integrated circuit from his notebook. (Courtesy of Texas Instruments.)

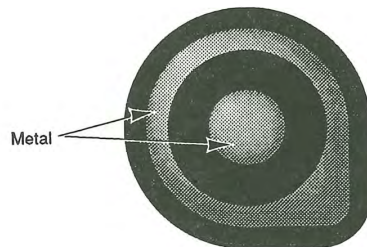


Figure 1.7 Horni "teardrop" transistor.

the needs of that era, but contained the seeds for the miniaturization and cost-effective manufacturing techniques that still drive the industry. Kilby and Noyce shared the patent for the integrated circuit.

Starting with the simple Kilby IC, the industry has never looked back. Each year the number of devices in a single circuit has increased. In 1964 Gordon Moore, a founder of Intel, predicted that integrated circuit density would double every year. This prediction became known as Moore's law and has proven very accurate (Fig. 1.8).

The industry keeps track of these integration levels by a scale (Fig. 1.9) ranging from small-scale integration (SSI) to ultra-large-scale integration (ULSI). ULSI chips are sometimes referred to as very very large-scale integration ((V)VLSI). The popular press calls these newest products *megachips*.

In addition to the integration scale, memory circuits are identified by the number of memory bits contained in the circuit (e.g., a four-meg memory chip can store four million bits of memory). Logic circuits are often rated by their number of gates. A *gate* is the basic operational component of a logic circuit.

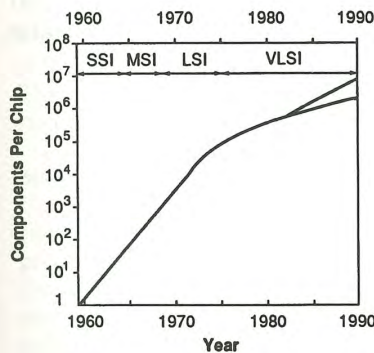


Figure 1.8 Growth in IC density [After Moore, "VLSI, What Does the Future Hold?", Electronic News, 42(14) (1980).]

Level	Abbreviation	# Components per Chip
Small Scale Integration	SSI	2 - 50
Medium Scale Integration	MSI	50 - 5000
Large Scale Integration	LSI	5000 - 100,000
Very Large Scale Integration	VLSI	Over 100,000 - 1,000,000
Ultra Large Scale Integration	ULSI	> 1,000,000

Figure 1.9 IC integration table.

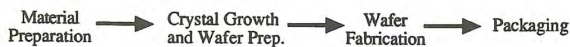


Figure 1.10 Stages of semiconductor manufacturing.

Stages of Manufacturing

Solid-state devices are manufactured in four distinct stages (Fig. 1.10). They are material preparation, crystal growth and wafer preparation, wafer fabrication, and packaging.

In the first stage, material preparation (see Chap. 2), the raw semiconducting materials are mined and purified to meet semiconductor standards. For silicon, the starting material is sand (Fig. 1.11).

In stage two the material is formed into wafers (Fig. 1.12) in a process called crystal growth and wafer preparation (Chap. 3). Wafers (also called slices) are thin disks of the semiconducting material. This stage is performed by either a separate department within the semiconductor company or vendors outside the company. Figure 1.11 shows the conversion of sand to semiconductor grade silicon. The industry also makes devices and circuits from germanium and compounds of different semiconductor materials.

It is in stage three, wafer fabrication, that the devices or integrated circuits are actually formed in and on the wafer surface (Fig. 1.13). Up to several thousand identical devices are formed on each wafer,

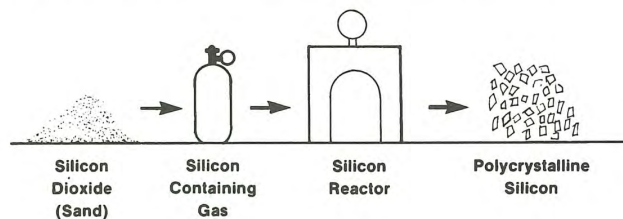


Figure 1.11 Conversion of silicon dioxide to semiconductor grade silicon.

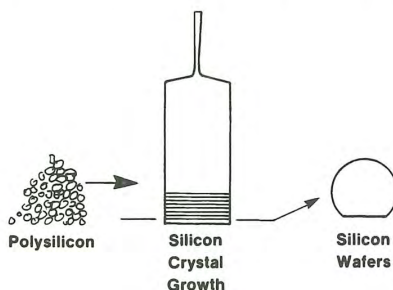


Figure 1.12 Crystal growth and wafer preparation.

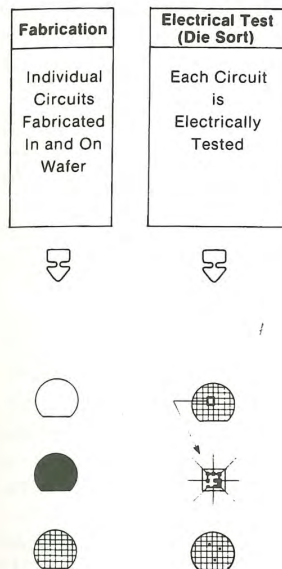


Figure 1.13 Wafer fabrication (and electrical test).

although 200 to 300 are more common. The area on the wafer occupied by the discrete device or integrated circuit is called a chip or die. The wafer fabrication process is also called *fabrication*, *fab*, *chip fabrication*, or *microchip fabrication*.

Chips are fabricated by both *merchant* and *captive* producers. Merchant suppliers manufacture chips to sell on the open market. Captive producers, whose final products are computers, communications systems, and the like, produce chips in-house for their own products. Some firms produce chips for in-house use and also sell on the open market, and others produce specialty chips in-house and buy others on the open market. During the 1980s a greater percentage of chips being fabricated in captive fab areas became the trend. Also, during the 1980s a number of firms started supplying chips but contracted out the wafer fabrication. They are known as *fabless chip suppliers*.

After wafer fabrication, the chips on the wafer are complete but untested. Next comes an electrical test (called *wafer sort*) of every chip to identify those that meet customer specifications. Wafer sort may be the last step in wafer fabrication or the first step in the *packaging* process. Packaging (Fig. 1.14) is the series of processes that separates the wafer into individual die and places them into protective packages.

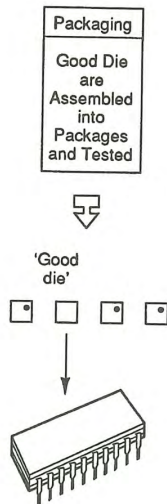


Figure 1.14 Packaging stage.

A protective chip package is necessary to protect the chip from contamination and abuse and to provide a durable and substantial electrical lead system, which will connect the chip onto a printed circuit board or directly into an electronic product. Packaging takes place in a different department of the semiconductor company and quite often in an offshore plant.

While most chips are packaged in individual packages, a growing percentage are being incorporated into hybrid circuits or multichip modules (MCMs) or mounted directly on printed circuit boards (chip-on-board or COB). An integrated circuit comprises entirely solid-state devices contained on one chip. A hybrid circuit combines semiconductor devices (discretes and ICs), with thick or thin film resistors and conductors, and other electrical components on a ceramic substrate. These techniques are explained in Chap. 18.

Process and Product Trends

Since 1947, the semiconductor industry has seen the continuous development of new and improved processes. These process improvements have led to more highly integrated and reliable circuits that have, in turn, fueled the continuing electronics revolution. These process improvements fall into two broad categories, process and structural. Manufacturing processes and price structure improvements result in the fabrication of the devices and circuits in ever

high
new
trol
T
pan
the
abo
(0.5
larg
give
A
devi
be p
of in
has
proc
cond
show
Pr
wafe
(I/O,
Indu
show
An
hood

Average design rule (μm)
10
1
0.1
15

Figure 1
"Silicon

higher density, quantity, and reliability. Structural improvements are new device designs that yield greater circuit performance, power control, and reliability.

The advancement of chip density from SSI to ULSI has been accompanied by two trends, increasing chip size and decreasing the size of the individual circuit components. Discrete and SSI chips average about 100 mils (0.1 in) on a side. ULSI chips are about 500–750 mil (0.5–0.75 in) per side. Increasing die size also pushed the drive to larger wafer diameters. The 1 in diameter wafers of the 1960s have given way to 200- and 300-mm (8- and 12-in) wafers.

Although integrated circuit chip size has grown as more and more devices are crowded into the circuits, the higher densities would not be possible without a commensurate decrease in the size of each part of individual devices, known as the device *feature size*. This decrease has been brought about by dramatic improvements in the imaging process, known as *photolithography*, and the trend to multilayers of conductors. The decrease in feature size over a twenty-year span is shown in Fig. 1.15.

Projections for these interdependent factors—feature size, die size, wafer diameter, number of interconnect levels, and package pin count (I/O, for in/out connections)—have been made by the Semiconductor Industry Association (SIA) through the year 2007. Their “road map” is shown in Fig. 1.16.

An analogy used to explain these trends is the layout of a neighborhood of single family homes. The density of the neighborhood is a func-

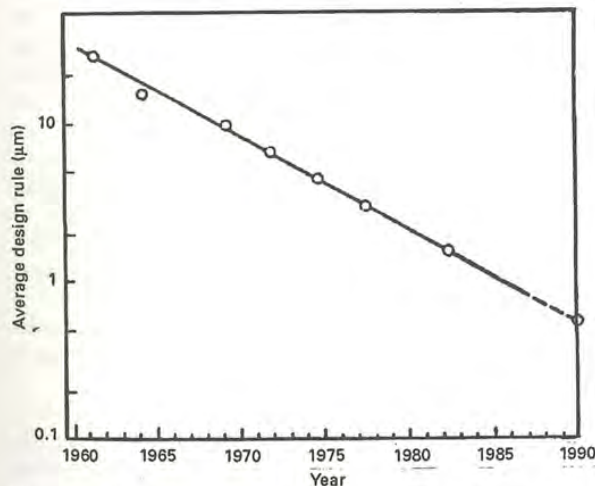


Figure 1.15 Decreasing image feature size. (After Wolf and Tauber, “Silicon Processing for the VLSI Era.”)

DRAM Production	1995	1998	2001	2004	2007
Feature Size (nm)	350	250	180	120	100
Bits (DRAM)	64M	256M	1G	4G	16G
Processing Cost (\$/cm²)	\$3.90	\$3.80	\$3.70	\$3.60	\$3.50
Chip Size (mm²)					
Logic	400	600	800	1000	1250
DRAM	200	320	500	700	1000
Defect Density (1/cm²)	0.05	0.03	0.01	0.004	0.002
Logic Interconnect Levels	4-5	5	5-6	6	6-7

Figure 1.16 SIA lithography roadmap. (Courtesy Semiconductor Industry Association)

tion of the house size, lot size, and width of the streets. Increasing the density can be accomplished by building more houses, which results in a larger neighborhood. Or the houses can be put closer together by reducing the lot and street size. Another possibility is to reduce the size of the individual houses. However, at some point the streets cannot be reduced anymore or they won't be wide enough for autos, and at some point the houses cannot be reduced in size and still function as dwelling units.

The same types of options take place in chip design. The individual parts of the devices can be reduced in size (houses) and moved closer together (lot and street size). A denser chip is the result.

There are several benefits to the reduction of feature size and the attendant increase in circuit density. At the circuit performance level there is an increase in circuit speed. With shorter distances to travel and with the individual devices occupying less space, information can be put into and gotten out of the chip in less time. Anyone who has waited for a personal computer to perform a simple operation can appreciate the effect of faster performance. These same density improvements result in a chip or circuit that requires less power to operate. The small power station required to run the ENIAC has given way to powerful laptop computers that run on a set of batteries.

A companion trend has been an increase in die size. The benefits of increasing chip density also apply to larger circuits. Making the individual chip components smaller and closer together *and* creating a larger chip yields faster and more powerful circuits.

Perhaps the most significant effect of these process and product improvements is the cost of the chips. Figure 1.17 shows the year-by-year drop in memory chips. The curves are typical for any maturing

.035μ
.025μ
.015μ
19μ

produ
manu
becom
perfor
mega-
1.18 a
cost, h
to sol
carrie
amou
Consi
achiev

26
21
16
11
6
1
Transistor cost (10⁻⁶\$)

Figure 1.17 Transistor cost and chip size. (Courtesy Semiconductor Industry Association)

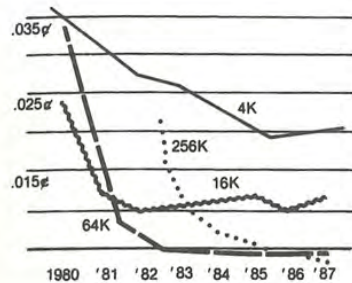


Figure 1.17 Price of chips per bit of memory.

product. Prices start high, but as the technology is mastered and manufacturing efficiencies increase, the prices drop and eventually become stable. These chip prices have constantly declined even as the performance of the chips has increased. However, the economics of mega-chip manufacturing may actually drive the cost of chips up (Fig. 1.18 and Chap. 15). The two factors, increased performance and lower cost, have driven the switch from conventional electronic components to solid-state devices. An example is the automobile. A modern auto carries hundreds of dollars of on-board computing power. That same amount of power would have cost some \$3000 in a 1970 auto. Considering that a 1970s car cost about \$3000, this is a remarkable achievement. Even more impressive is the personal computer. Today,

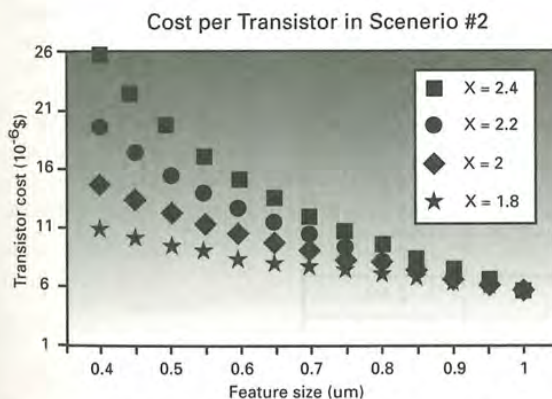


Figure 1.18 Rising cost of transistors with decreasing feature size and decreasing manufacturing efficiency (X). (Courtesy Semiconductor International, June 1994)

CHIP USES	1996
Computer	48.0%
Consumer	21.2%
Telecom	14.7%
Industrial	9.8%
Automotive	4.4%
Military	1.9%
Total	100.0%

Figure 1.19 Semiconductor chip uses. (Courtesy In-Stat-1995 SEMI ISS seminar)

a moderate priced desktop computer can deliver the same power as an IBM mainframe manufactured in 1970.

The availability of cheap chips has led to their inclusion in many different products. Major industry use of chips is shown in Fig. 1.19.

To offset the decline in chip prices, the semiconductor industry has continually had to make dramatic improvements in processing technology, including vendor-supplied materials and equipment. The fabrication processes have become more numerous as the use of automation is becoming an increasing requirement in many fields (see Chap. 15).

Overall, the semiconductor industry has experienced continual growth. From its birth in the 1950s it is now approaching worldwide sales of \$2 trillion a year, supported by a supplier industry of over \$1 trillion. Millions of chips are supplied by factories located throughout the world. Interestingly, even as the industry shows signs of maturing it is still growing faster than other "mature" industries, indicating that microchips are nowhere near the technical and performance limits (Fig. 1.20).

KEY INDUSTRY	% Growth*
Automobiles	1.3%
Pharmaceuticals	1.5%
Aircraft	2.1%
Computers/Peripherals	2.5%
GDP	3.0%
Telecom-Equipment	3.7%
Semiconductors/Elec. Comp.	9.3%

* Compound Annual Growth Rate 1987-1994

Figure 1.20 Key industry growth rates. (Courtesy U.S. Industrial Outlook 1994, SEMI ISS 1994)

Year

DRAM
Feature
Volume

Figure 1.21

An exa
indicates
can be s
(DRAM)

Five Dec

The histo
ments ar
al refine
The first
industry

The devel

While th
ognized
realized
work and
today.

The Be
tor becau
wires ont
quickly g

Emitter p

Figure 1.22

Year	1993	1996	1999	2002	2005
DRAM Capacity	16Mb	64Mb	256Mb	1Gb	4Gb
Feature Size (micron)	0.5	0.35	0.25	0.18	0.08
Volumes of Enc. Britannica	1/4 Vol	1 Vol.	4 Vol.	16 Vol.	2 Sets

Figure 1.21 Future DRAM capacity. (Source: BusinessWeek, July 4, 1994)

An example of increasing chip power is shown in Fig. 1.21, which indicates the number of volumes of the Encyclopedia Britannica that can be stored on larger capacity dynamic random access memory (DRAM) chips.

Five Decades of Progress

The history of the semiconductor industry is one of continual developments and advances. Each new level of process expertise and material refinement has made possible newer and more powerful circuits. The first five decades of the semiconductor era have seen an emerging industry rise to world dominance.

The development decade (1951–1960)

While the tremendous advantages of solid-state electronics were recognized early, the advances possible from miniaturization were not realized until two decades later. During the 1950s engineers set to work and defined many of the basic processes and materials still used today.

The Bell transistor came to be known as the *point contact* transistor because the surface contacts were made by springing two thin wires onto the semiconductor surface. This system was unreliable and quickly gave way to the *alloy junction* transistor (Fig. 1.22). In this

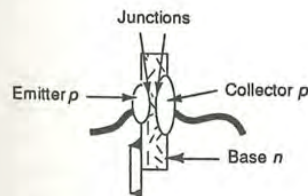


Figure 1.22 Alloy junction transistor.

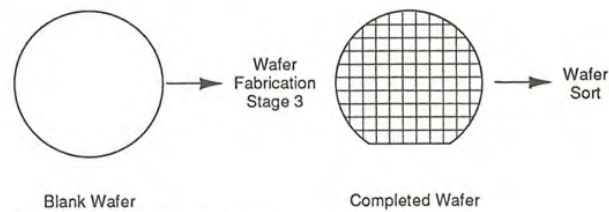


Figure 5.1 Wafer-fabrication stage.

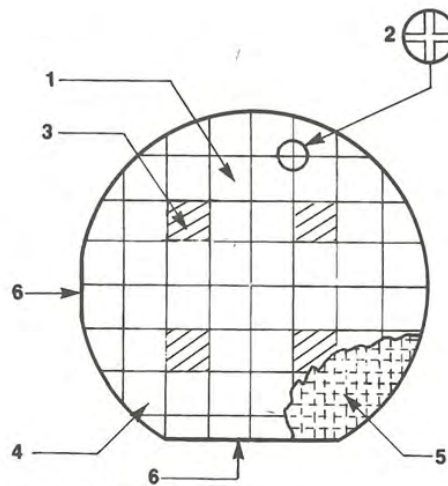


Figure 5.2 Wafer terminology.

Goal of Wafer Fabrication

The four stages of semiconductor manufacturing are materials preparation, crystal growth and wafer preparation, wafer fabrication, and packaging (see Chap. 1). The first two stages have been explored in Chap. 3. In this chapter the fundamentals of wafer fabrication are explained.

Wafer fabrication is the series of processes used to create the semiconductor devices in and on the wafer surface. The polished starting wafers come into fabrication with blank surfaces and exit with the surface covered with hundreds of completed chips (Fig. 5.1).

Wafer Terminology

A completed wafer is illustrated in Fig. 5.2. The regions of a wafer surface are:

1. Chip, die, circuit, microchip, or bar. All of these terms are used to identify the identical patterns covering the majority of the wafer surface.
2. Scribe lines, saw lines, streets, and avenues. These areas are small avenues of space between the chips used to separate the chips from each other. Generally the scribe lines are blank, but some companies place alignment targets (see photomasking) in them.
3. Engineering die, test die. These chips are different from the regular device or circuit die. They contain special devices and circuit elements that can be electrically tested during fabrication processing.
4. Edge die. The edges of the wafer contain partial die patterns. The partial die will not function. The number and area occupied by the edge die is a function of the chip size and the wafer diameter. One of the driving forces behind larger wafer diameters is to minimize the area occupied by the edge die.
5. Wafer crystal planes. The cutaway section illustrates the crystal structure of the wafer under the circuit layers. The diagram shows that the chip edges are oriented to the wafer crystal structure.
6. Wafer flats. The depicted wafer has a major and minor flat, indicating that it is a P-type $\langle 100 \rangle$ oriented wafer (see Chap 3. for the flat code).

Basic Wafer-Fabrication Operations

Wafer-fabrication areas around the world produce billions of chips with thousands of different functions and designs. However, there are a limited number of transistor types. Most are variations on either a bipolar or MOS structure (see Chap. 16). Additionally, the choices for wafer materials are few in number (as explained in Chap. 2). Even so, the variations in manufacturing processes are almost infinite, with each company exercising its own version of a standard process. But within all this process diversity there are only four (Fig. 5.3) basic operations performed on a wafer in the fabrication process. They are:

1. Layering
2. Patterning
3. Doping
4. Heat treatments

each major step, for contamination. Wafers that exceed the established allowable density are rejected.

Mask defects

A photomask or reticle is the source of the pattern that is transferred to the wafer surface in the patterning process. Defects on the mask or reticle end up on the wafer as defects or pattern distortions. There are three common defects that originate from masks or reticles. The first is contamination from dirt or stains on the clear part of the mask or reticle. In optical lithography, dirt can block the light and print onto the wafer as though it was an opaque part of the pattern. Second are cracks in the quartz. They too can block the patterning light or scatter the light, causing unwanted images and/or distorted images. Third are pattern distortions that may have occurred in the mask- or reticle-making process. These include pinholes or chrome spots, pattern extensions or missing parts, breaks in the pattern, or bridges between adjacent patterns (see Fig. 6.4).

Control of mask-generated defects is more critical for devices and circuits with smaller feature sizes, higher densities, and larger die sizes.

Wafer-Sort Yield Factors

After fabrication, the wafers go to the wafer-sort tester. During the test, each chip will be tested electrically for device specifications and functionality. Up to several hundred individual electrical tests may be performed on each circuit. While these tests measure the electrical performance of the device(s), they indirectly measure the precision and cleanliness of the fabrication processes. Because of natural process variations and undetected defects, the wafer may have passed all the in-process checks and still have many chips that do not function.

Since wafer sort is a comprehensive test, many factors influence the yield. They are:

1. Wafer diameter
2. Die size (area)
3. Number of processing steps
4. Circuit density
5. Defect density
6. Crystal defect density
7. Process cycle time

Figure 6
trusion
July 19

Wafer
The s
tion o
that
wafer
lines
ductio
ter wa

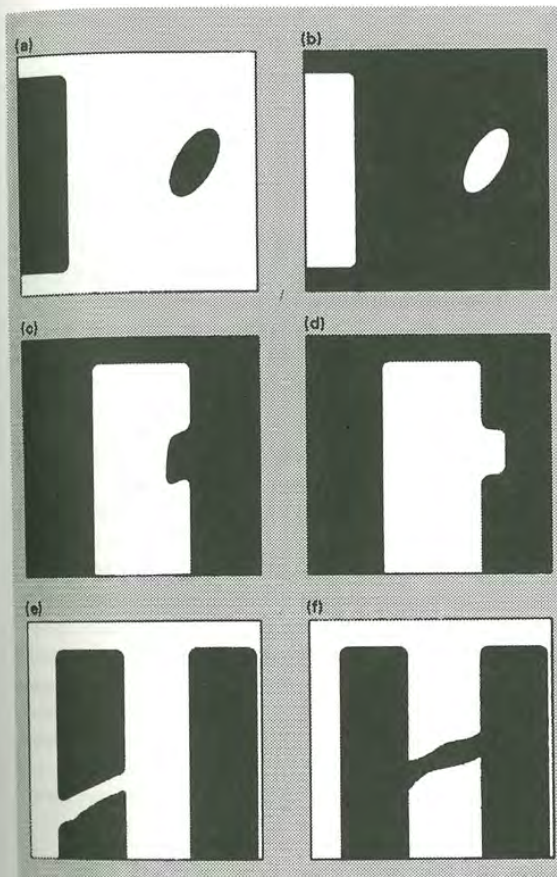


Figure 6.4 Mask defects: (a) spot; (b) hole; (c) inclusion; (d) protrusion; (e) break; (f) bridge. (Source: *Solid State Technology*, July 1993, page 95).

Wafer diameter and edge die

The semiconductor industry went to round wafers with the introduction of silicon. The first wafers were less than 1 in in diameter. Since that time there has been a regular progression to larger-diameter wafers, with 150 mm (6 in) being the standard of VLSI fabrication lines in the late 1980s and 200-mm wafers being developed for production use in the 1990s. Projections call for 300- and 400-mm diameter wafers early in the 21st century.¹

The move to larger-diameter wafers has been driven by production efficiency, increasing die sizes, and the influence on the wafer-sort yield. Production efficiency is easily understood when one considers that there is an incremental increase in the cost of processing a larger-diameter wafer while the number of available whole chips on the wafer can increase substantially as illustrated in Fig. 6.5.

The effect of increasing the wafer diameter also has positive effects on the wafer-sort yield. Fig. 6.6 shows two wafers of the same diameter but with different die sizes. Note that the smaller-diameter wafer has a very large proportion of its surface covered with partial die, die that cannot function. The larger-diameter wafer with its greater number and percentage of whole die will, if all other factors are equal, have a higher wafer-sort yield.

Wafer diameter and die size

Another driving force for larger-diameter wafers is the trend to larger die sizes. As shown in Fig. 6.6, increasing the die size *without* increasing the wafer diameter also results in a wafer surface with a smaller percentage of whole die.

Maintaining a decent wafer-sort yield as the die size increases requires increasing the wafer diameter. Figure 6.7 lists the number of

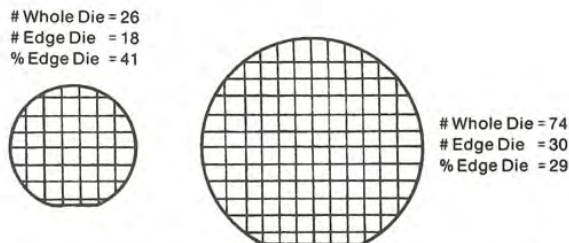


Figure 6.5 Effect of larger wafer diameter on percentage of partial die.

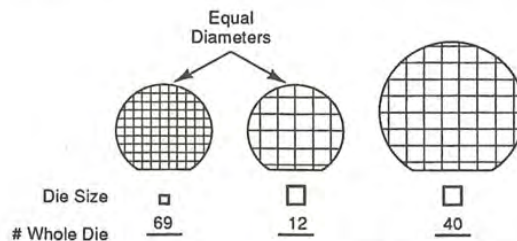


Figure 6.6 Effect of processing larger die on larger wafers.

Die Size (mils)	Wafer Diameter (mm)			
	100	125	150	200
200x200	283	452	660	1194
300x300	126	201	293	531
400x400	71	113	165	298
500x500	45	72	106	191

Figure 6.7 Die size versus number of die on wafer.

various size chips that will fit on different size wafers. The bottom line is that larger-diameter wafers are more cost-effective.

Wafer diameter and crystal defects

In Chap. 3 the concept of a crystal dislocation was introduced. A crystal dislocation is a point defect *in the wafer* that comes from a local discontinuity of the crystal structure. Dislocations exist throughout the crystal structure and, like contamination and process defect density, affect the wafer-sort yield.

Dislocations also are generated during the fabrication process. They generate (or nucleate) at sites where there are chips and abrasions of the edge of the wafer. These chips and abrasions come from poor handling techniques and automatic handling equipment. The abraded area causes a crystal dislocation. Unfortunately, the dislocation is propagated into the center of the wafer (Fig. 6.8) during subsequent heat treatments, such as oxidations and diffusions. The length

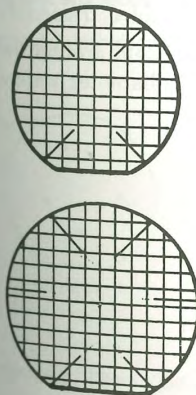


Figure 6.8 Effect of dislocations on wafer-sort yield for different wafer diameters.

of the dislocation line into the interior of the wafer is a function of the thermal history of the wafer. Consequently, wafers receiving more process steps and/or more heating steps will have more and longer dislocation lines. One obvious solution to the problem is larger-diameter wafers. This solution leaves a larger number of unaffected die in the center of the wafer.

Wafer diameter and process variations

The process variations discussed under "Wafer-Fabrication Yield Limiters" affect the wafer-sort yield. In the fabrication area, process variations are detected by sampling inspection and measurement techniques. The nature of inspection sampling is that not all of the variations and defects are detected so that wafers are passed on with some number of problems. These problems show up at wafer sort as failed devices.

Process variations occur at a higher rate around the edge of the wafer. In the high-temperature processes performed in tube furnaces, there is always some temperature nonuniformity across the wafers. The change in temperature results in uniformity differences on the wafer. Variations occur more at the wafer's outer edges where heating and cooling occur at a faster rate. Another contributor to this wafer edge phenomenon is contamination and physical abuse of the wafer layers that emanate from handling and touching the wafers on their edges. Also, in the patterning process there can be feature size uniformity problems in the mask-driven processes (full-mask projection, proximity, and contact exposure). The nature of the light systems is such that the center will be of higher uniformity than the outside edges. In the reticle-driven masking processes (steppers) there is a smaller area of exposure (one or several die), which reduces the image variations across the wafer.

All of these problems result in a lower wafer-sort yield around the edge of the wafer, as illustrated in Fig. 6.9. Larger-diameter wafers

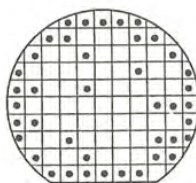


Figure 6.9 Typical location of functioning die after wafer sort.

• Nonfunctioning Die

help to maintain wafer-sort yields by having a larger area of unaffected die in the center of the wafer.

Die area and defect density

The die size also affects wafer-sort yield relative to the defect density on the wafer surface. The relationship is illustrated in Fig. 6.10. In Fig. 6.10a, a wafer is shown with five defects and no die pattern. This situation illustrates a background defect density produced by all the fabrication area factors, regardless of the die size, device type, process control requirements, and so on. The wafers in Figs. 6.10b and c illustrate the effect of this background defect density on the wafer-sort yield for two different die sizes. The larger the die size for a given defect density, the lower the yield.

Circuit density and defect density

The defects on the wafer surface result in die failures by causing a malfunction of some part of the die. Some of the defects are located in nonsensitive parts of the die and do not cause a failure. However, the trend is to higher levels of circuit integration, which has come about because of smaller feature size and higher-density die components. The result of these trends is a higher probability that any given defect will be in an active part of the circuit, thus lowering the wafer-sort yield as illustrated in Fig. 6.11.

Number of process steps

The number of process steps was earlier shown to be a limiter of the cum yield. The more steps there are, the more opportunities there are to break or otherwise damage a wafer. This effect also influences the wafer-sort yield. As the number of process steps increases, the background defect density increases unless procedures are implemented to

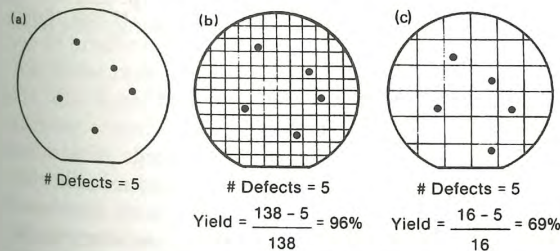


Figure 6.10 Effect of defects of sort yield for different die sizes.

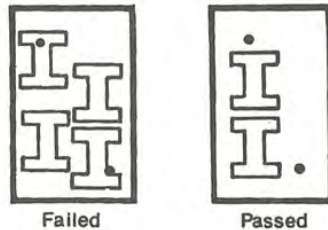


Figure 6.11 "Killer" defects and nonfatal defects.

lower it. A higher background defect density affects more chips, thus lowering the wafer-sort yield.

Feature size and defect size

Smaller feature sizes make maintaining an acceptable sort yield difficult because of two major factors. First, the smaller images are more difficult to print (see "Mask Defects" and Chap. 8). Second, the smaller images are vulnerable to ever-smaller defect sizes as well as the overall defect density. The 10 to 1 ratio rule of minimum feature size to allowable defect size has been discussed. One assessment is that at a defect density of 1 defect per cm^2 a circuit with a $0.35\text{-}\mu\text{m}$ feature size will have a wafer-sort yield 10 percent less than a $0.5\text{-}\mu\text{m}$ circuit processed under the same conditions.²

Process cycle time

The time that the wafers are actually being processed can be measured in days. But because of queuing at the process stations and temporary slowdowns due to process problems, the wafer often stays in the fabrication area for several weeks. The longer the wafer stays in-process, the more opportunity there is for contamination, which in turn will lower wafer-sort yield.

The move to just-in-time manufacturing (see Chap. 15) is one attempt to increase yields and decrease the manufacturing costs associated with high in-line inventories.

Wafer-sort yield formulas

The ability to understand and predict wafer-sort yields with some accuracy is essential to the operation of a profitable and reliable chip supplier. Over the years, a number of models have been developed that relate process, defect densities, and chip size parameters to the wafer-sort yield. Four yield model formulas are shown in Fig. 6.12.

(a) Exponential $Y = \frac{1}{e^{AD}}$	(c) Murphy $Y = \left(\frac{1 - e^{DA}}{DA} \right)^2$
(b) Seeds $Y = \frac{1}{(1 + DA)^n} \left(\frac{r - \sqrt{A}}{\sqrt{A}} \right)^2$	(d) "New" $Y = \left(\frac{N_{tr} d_d D_1}{\lambda^{p-2}} \right)$

where Y = percent functioning die A = die area D = defect density n = number of masking steps r = radius of wafer	N_{tr} = number of transistors d_d = number of minimum feature size "squares" D_1 = defect characterization parameter p = defect characterization parameter λ = minimum feature size
---	--

Figure 6.12 Wafer-sort yield models.

Each relates different parameters to the wafer-sort yield. As the chips get larger, the number of process steps increases, and the feature size decreases, the sensitivity to smaller defect sizes increases and thus more background defects become killer defects.

Exponential model. The exponential relationship (Fig. 6.12a) is the simplest and one of the first yield models³ developed. It is applicable for wafers that contain over 300 die and MSI circuits of lower densities. The model does not have a factor for edge die as does the Seeds model. The exponential model is a valuable learning tool because it clearly shows the primary relationships among die area, defect density, and wafer-sort yield. To understand the relationship in this mathematical model, remember that e is a constant with a value of 2.718. The exponential model shows that any increase in die area or defect density will increase the value of the divisor on the right-hand side of the equation and drive the wafer-sort yield down.

Seeds model. The Seeds model adds two additional parameters to the job of predicting wafer-sort yield, the number of process steps (n) and the wafer radius–die size relationship r (r and a). In most yield models, the factor for processing steps (n) is actually the number of patterning steps. Experience has proved that the patterning steps contribute the greatest number of point defects and therefore have a direct bearing on sort yield.

Murphy model. Perhaps the most widely used yield model is the Murphy relationship (Fig. 6.12c). While not including a factor for the

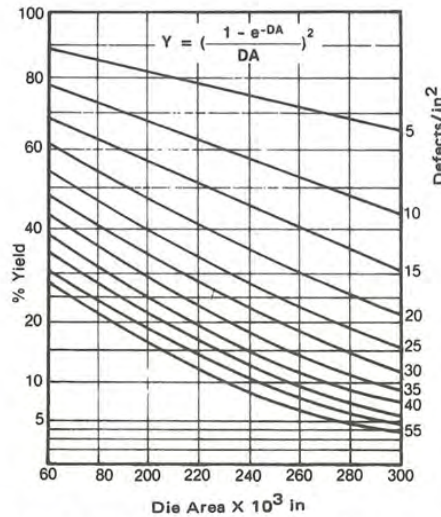


Figure 6.13 Murphy's yield model, showing die yield as a function of die size defect density.

number of processing steps, it has proved a good predictor for LSI-, VLSI-, and ULSI-level circuits. Figure 6.13 shows a graphical solution of the Murphy model, where the relationships of sort yield, defect density, and die size are clearly shown.

High-density, small-feature-size model. A newer model (Fig. 6.12d) factors in the minimum feature size (λ), the area occupied by a transistor (d_d), and the design density (lower values of d_d). This model is complicated and several of the parameters had to be fine-tuned to the experimental data. (The reader is directed to reference 4 for a better understanding of the model.) But it illustrates the complexity of relating wafer-sort yield to the many sensitive factors in high-density, small-feature-size chips.⁴ Complex circuits rarely have comparable designs and processes vary from company to company as does the basic background defect density. These factors make the development of an accurate universal yield model difficult. Most chip companies have developed their own models that reflect their particular manufacturing processes and product designs.

The defect density used in all the models is not the same as a defect density determined by optical inspection of the wafer surface. The defect density that shows up in the yield models is all-inclusive; it includes contaminants and surface and crystal defects. Further, it predicts only the defects that destroy die: the "killer defects." Defects that

Overall Yield Formula			
(Wafer Fab Yield)	(Wafer Sort Yield)	(Packaging Yield)	= Overall Yield
$\frac{\# \text{ Wafers Out}}{\# \text{ Wafers Started}} \times \frac{\# \text{ Functioning (Good) Die}}{\# \text{ Die On Wafer(s)}} \times \frac{\# \text{ Packages Passing Final Test}}{\# \text{ Die Started Into Packaging}} = \text{Overall Yield}$			

Figure 6.15 Overall yield formula.

gives the percentage of shipped die as compared with the number of whole die on the starting wafer. It is an inclusive measurement of the success of the entire process.

Overall yields vary with several major factors. In Fig. 6.16 is a list of typical process yields and their calculated overall yield. In the first two columns are major process factors that influence the individual and overall yields.

First is the integration level of the particular circuit. The more highly integrated the circuit, the lower the expected yield in all categories. Higher integration levels assume a corresponding decrease in feature size. Column 2 lists the maturity of the manufacturing process. Process yields almost always follow an S curve pattern (Fig. 6.17) through the lifetime of the product in manufacturing. In the beginning, the yield rises rather slowly as the initial bugs are worked out of the process. This is followed by a period when the yields rise rapidly, eventually leveling off because of the limits imposed by the process maturity, including die size, integration level, circuit density, and defect density. If the product goes into a period of steadily declining volume, the causes may be older equipment and less engineering attention.

As the table in Fig. 6.16 shows, overall yields can vary from very low (maybe even zero for new or poorly designed products) to the 90

Integration Level	Product Maturity	Fab Yield %	Sort Yield %	Assembly Final Test %	Overall Yield %
ULSI	Mature	95	85	97	78
ULSI	Mid	88	65	92	53
ULSI	Introduction	65	35	70	16
LSI	Mature	98	95	98	91
Discrete	Mature	99	97	98	94

Figure 6.16 Typical yields for various products.

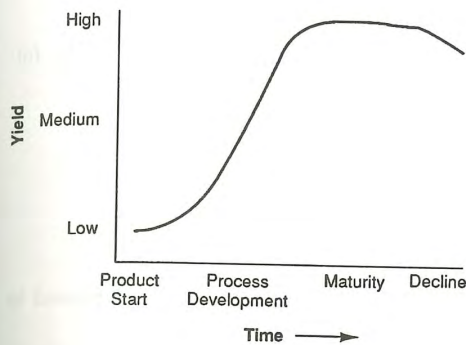


Figure 6.17 Yield changes with process maturity levels.

percent range for simpler or more mature products. Semiconductor producers consider their yield performance very proprietary since profit and production control are a direct function of the process yields.

An examination of the yield values in the table reveals that wafer sort is the lowest of the three yield points. This fact illustrates why yield-improvement programs are directed at the many factors that influence the wafer-sort yield. Improvement of wafer-sort yields has the biggest impact on productivity. The advent of larger and more complex chips (such as megabit memory chips) has shifted productivity improvements to include other factors including the cost of ownership of equipment (see Chap. 15). Successful competition in the megachip era will require wafer-sort yields in the 85–90 percent range.⁵

Key Concepts and Terms

Assembly and final test yields	Typical overall yields
Defect density	Wafer-fabrication yield factors
Murphy's yield formula	Wafer-sort yield factors
Overall yield factors	Wafer-sort yield formulas
Three major yield points	

Review Questions

1. Name the three yield measurement points in the manufacturing process.

surface end up in the new layer of oxide, away from the electrically active surface. Other contaminants are drawn up into the silicon dioxide film where they are less harmful to the devices. In the early days of MOS device processing, it was common to oxidize the wafers and then remove the oxide before further processing to rid the surface of unwanted mobile ionic contamination.

Doping barrier

In Chap. 5 doping was identified as one of the four basic fabrication operations. Doping requires creating holes in a surface layer through which specific dopants are introduced into the exposed wafer surface through diffusion or ion implantation. In silicon technology, the surface layer is most often silicon dioxide (Fig. 7.2). The silicon dioxide left on the wafer acts to block the dopant from reaching the silicon surface. All of the dopants used in silicon technology have a very slow rate of movement in silicon dioxide as compared to silicon. While the dopants penetrate to the required depth in the exposed silicon, they only penetrate a short distance into the silicon dioxide surface. It only takes a relatively thin silicon dioxide layer to block the dopants from reaching the silicon surface.

Another factor favoring the use of silicon dioxide is that it possesses a coefficient of thermal expansion similar to that of silicon. In high-temperature processes such as oxidation and diffusion doping, the wafer expands and contracts as it is heated and cooled. The silicon dioxide expands and contracts at close to the same rate as silicon, which means that the wafer will not warp during the heating and cooling.

Surface dielectric

Silicon dioxide is classified as a dielectric. This means that under normal circumstances it does not conduct electricity. When dielectrics are used in electrical circuits or devices they are referred to as *insulators*. Acting as an insulator is an important role of silicon dioxide layers. Figure 7.3 shows a cross section of a wafer with a conductive layer of metal on top of a layer of silicon dioxide. The oxide prevents shorting

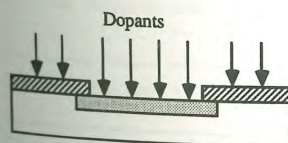


Figure 7.2 Silicon dioxide layer as dopant barrier.

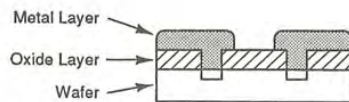


Figure 7.3 Dielectric use of silicon dioxide layer.

of the metal layer to the underlying metal just as the insulation on an electric cord prevents the wires from shorting. In this capacity the oxide must be continuous, that is, having no holes or voids.

The oxide must also be thick enough to prevent a phenomenon known as *induction*. Induction can occur when the separating layer of oxide is thin enough to allow an electrical charge in a metal layer to cause a buildup of charge in the wafer surface. The surface charge can cause shorting and other unwanted electrical effects. A thick enough layer will prevent an induced charge in the wafer surface. An oxide layer thick enough to prevent induction from the metal layers is called the *field oxide*.

Device dielectric

At the other end of induction phenomena is MOS technology. In a MOS transistor, a thin layer of silicon dioxide is grown in the gate region (Fig. 7.4). The oxide functions as a dielectric whose thickness is chosen *specifically* to allow induction of a charge in the gate region under the oxide. The gate is the part that controls the flow of current through the device (see Chap. 16). The dominance of MOS technology for ultra-large-scale integrated (ULSI) circuits has made the formation of gate regions a prime focus of process development and concern. Thermally grown oxides are also used as the dielectric layer in *capacitors* formed between the silicon wafer and a surface conduction layer (Fig. 7.5).

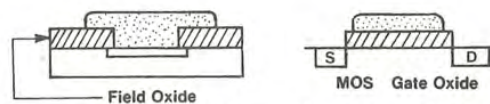


Figure 7.4 Silicon dioxide as field oxide and in MOS gate.

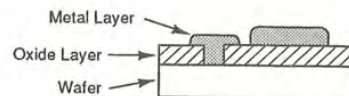


Figure 7.5 Silicon dioxide layer in solid-state capacitor.

Silicon dioxide dielectric layers are also used in structures with two or more metallization layers. In this application, the silicon dioxide layers are deposited with *chemical vapor deposition* (CVD) techniques rather than thermal oxidation (see Chap. 12).

Thermal Oxidation Mechanism

The silicon dioxide layers used in silicon-based devices vary in thickness. At the thin end of the scale are advanced MOS gate oxides, in the 35 to 80 Å range.¹ At the thick end are field oxides. Figure 7.6 lists the thickness ranges for the major uses.

Most of the silicon dioxide layers are grown on the silicon surface by thermal oxidation. Thermal oxide growth is a simple chemical reaction as shown in Fig. 7.7. This reaction takes place even at room temperature. However, an elevated temperature is required to achieve quality oxides in reasonable process times for practical use in circuits and devices. Oxidation temperatures are between 900 and 1200°C.

Although the formula in Fig. 7.7 shows the reaction of silicon with oxygen, it does not illustrate the growth mechanism of the oxide. To understand the growth mechanism, consider a wafer placed in a heated chamber and exposed to oxygen gas (Fig. 7.8a). Initially, the oxygen atoms combine readily with the silicon atoms. This stage is called *linear* because the oxide grows in equal amounts for each unit of time (Fig. 7.8b). After approximately 500 angstroms (Å) of oxide is grown, a limit is imposed on the linear growth rate. [An angstrom is one ten-thousandth of a micron (μm); in other words, there are 10,000 Å in 1 μm.]

Silicon Dioxide Thickness, Å	Application
60–100	Tunneling Gates
150–500	Gate Oxides, Capacitor Dielectrics
200–500	LOCOS Pad Oxide
2,000–5,000	Masking Oxides, Surface Passivation
3,000–10,000	Field Oxides

Figure 7.6 Silicon dioxide thicknesses chart.

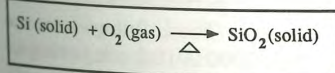


Figure 7.7 Reaction of silicon and oxygen to form silicon dioxide.

offer the advantage of higher deposition rates without increased temperatures.¹⁹

SOS and SOI

These two acronyms stand for *silicon on sapphire* and *silicon on insulator*. Both refer to the deposition of silicon on a nonsemiconductor surface. The need for such structures came about from the limits placed on some MOS devices by the presence of a semiconducting substrate under the active device. These problems are resolved by forming a silicon layer on an insulating substrate. The first substrate used for this purpose was sapphire (SOS). As different substrates were investigated, the term was expanded to the more general silicon on insulator (SOI).

One technique is a direct deposition on the substrate followed by a recrystallization process (laser heating, strip heaters, oxygen implantation) to create a usable film.²⁰ Another approach is a selective deposition through holes in a surface oxide with an overgrowth to form the continuous film.

Another SOI method is called SIMOX. In this process the top layer of a wafer is converted to oxide with a heavy oxygen implant. An epitaxial layer is grown on top of the oxide. There has been some exploration of *bonded wafers* as well. In this approach, two wafers are bonded together followed by the thinning (grinding and polishing) of one wafer to a thin device layer thickness.²¹

Insulators and Dielectrics

CVD is the favored method of depositing films that will function in the device or circuit as insulators or dielectrics. The two films in widespread use are silicon dioxide and silicon nitride. In general, the two films find a multiplicity of uses in device and circuit designs. While they have processing and quality differences, they must meet the same general requirements as other deposited films.

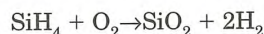
Silicon dioxide. Deposited silicon dioxide films are best known from their long-term use as a final passivation layer covering the completed wafer. In this role, they provide physical and chemical protection to the underlying circuit devices and components. Deposited silicon dioxide films used as a protective top layer are known by the proprietary terms *Vapox*, *Pyrox*, or *Silox*. Vapox (vapor-deposited oxide) is a term coined by Fairchild engineers. Pyrox stands for pyrolitic oxide. Silox is a registered trademark of Applied Materials, Inc. Sometimes the layer is sim-

ply called a *glass*. This protective role has expanded, and deposited silicon oxide layers are used as interdielectric layers in multimetallization schemes, as insulation between polysilicon and metallization layers, as doping barriers, as diffusion sources, and as isolation regions.

Silicon dioxide has become a major part of silicon gate structures. These structures include gate stacks, which consist of thermal oxide or silicon dioxide, or oxinitride, or silicon dioxide (TEOS deposited), and various silicon dioxide fillers for plugs in multimetal designs.²²

CVD-deposited silicon dioxide films vary in structure and stoichiometry from thermally grown oxides. Depending on the deposition temperature, deposited oxides will have a lower density and different mechanical properties, such as index of refraction, resistance to cracking, dielectric strength, and etch rate. These factors are highly affected by the addition of dopants to the film. In many processes the deposited film will receive a high-temperature anneal, a process called *densification*. After the densification, the deposited silicon dioxide film is close to the structure and properties of a thermal oxide.

The need for a low-temperature-deposited SiO_2 was dictated by the unacceptable alloying of aluminum and silicon at temperatures above 450°C . The early deposition process used was a horizontal conduction-heated APCVD system from silane and oxygen by the reaction



This process produced films that were of unacceptable quality for use in advanced device designs and on larger wafers due to the poorer film quality produced by the 450°C deposition temperature.

The development of LPCVD systems made possible higher-quality films, especially for the factors of step coverage and lower stress. LPCVD processes are the preferred deposition techniques from both quality and productivity considerations. High-temperature (900°C) LPCVD of silicon dioxide is performed with a dichlorosilane reaction with nitrous oxide:



Tetraethyl orthosilicate (TEOS). By far, most silicon dioxide is deposited from $\text{Si}(\text{OC}_2\text{H}_5)_4$ sources. The source is known as tetraethyl orthosilicate or TEOS. TEOS history goes back to the 1960s. Early systems relied on the simple pyrolysis of the TEOS in the 750°C range. Current depositions are based on the hot-wall LPCVD systems established in the 1970s, with temperatures of 400°C and above. TEOS sources used with plasma-assistance (PECVD or PETEOS) allowed deposition at temperatures below 400°C .²³ This process faces limits on conformal coverage of high-aspect ratio patterns in $0.5\text{-}\mu\text{m}$

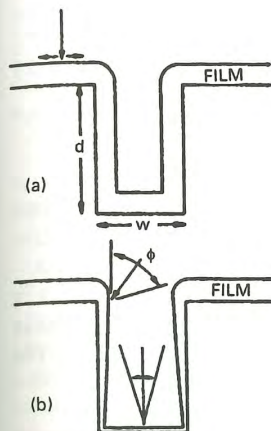


Figure 12.38 Step coverage. (a) good step coverage; (b) nonconformal coverage.

devices (Fig. 12.38). Step coverage is improved by the addition of ozone (O_3) to the gas stream.²⁴

Another option is the reaction of silane with nitrous oxide in an argon plasma:



Doped silicon dioxide. Silicon dioxide layers are doped to improve their protective characteristics and flow properties, or for use as dopant sources. The earliest dopant used with deposited oxides was phosphorus. The phosphorus source is phosphine (PH_3) gas added to the deposition gas stream. The resultant glass is called phosphorus silicate glass or PSG. Within the glass the phosphorus is in the form of phosphorus pentoxide (P_2O_5), making the glass a dual compound or, more correctly, a binary glass.

The role of the phosphorus is threefold. The added dopant increases the moisture-barrier property of the glass. Mobile ionic contaminants become attached to the phosphorus and are prevented from traveling into the wafer surface. This action is called *gettering*. The third result is an increase of the flow characteristics (Fig. 12.39) which aid the planarization of the glass surface after a heating step in the 1000°C range. The phosphorus content is limited to about 8 weight by percent. Above this level the glass becomes hygroscopic and attracts moisture. The moisture can react with the phosphorus, form phosphoric acid, and attack underlying metal lines.

Boron is often added to the glass from a diborane (B_2H_6) source. The purpose of the boron is to also aid the flow characteristics

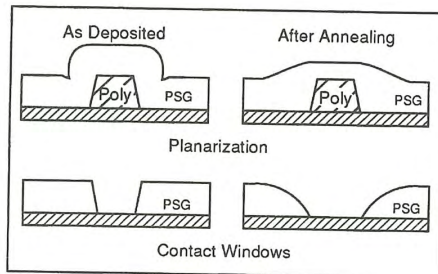


Figure 12.39 Planarization of surface by flowing glass.

(Fig. 12.39). The resultant glass is called a borosilicate glass (BSG). The boron and phosphorus are often used together in the glass. The result is referred to as BPSG (borophosphorus silicate glass).

Silicon nitride. Silicon nitride is a replacement for silicon dioxide uses, especially for top layer protection. Silicon nitride is harder (which provides better scratch protection), is a better moisture and sodium barrier (without doping), has a higher dielectric strength, and resists oxidation. The latter property has led to its use in the local oxidation of silicon (LOCOS) for isolation purposes. Figure 12.40 illustrates the process, where patterned islands of silicon nitride prevent oxidation under the islands. After thermal oxidation and removal of the nitride, there are wafer surface regions ready for device formation, separated by isolating regions of oxide. A disadvan-

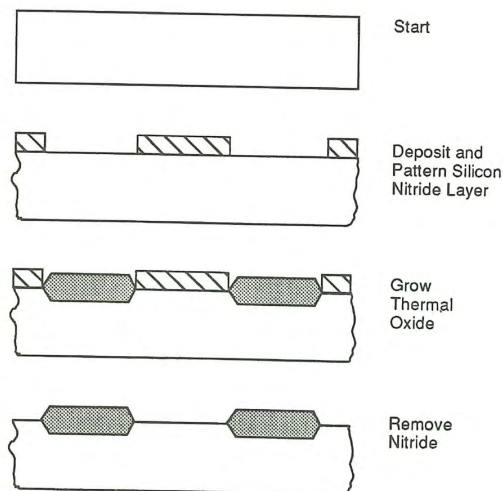


Figure 12.40 LOCOS process.

Silane
Dichloro

Figure 12.4

tage of s
and is m
with the

An ear
of a low-
perature
nitride fr
with the
reactors b
minum m
use of dif
ammonia

Conductor

The traditi
deposited
Perhaps th
Polysilicon
doping elem
ductors. H
deposited re
ductor mate
explained in

Key Concept

Amorphous
APCVD
Barrel system
BPSG
BSG
CVD
Deposited Si_3N_4
Deposited SiO_2
Epitaxial
Induction heating
LPCVD

will need to have one as a condition of getting the contract. The anti-static program is based on operators wearing grounding straps and nonstatic smocks; the use of antistatic carrier materials; moving work by lifting rather than sliding; and grounded equipment, work surfaces, and floor mats. Static is also reduced by the placement of ionizers in nitrogen and air blow-off guns (Fig. 18.7) and in the path of air coming out of HEPA filters.

Basic processes

In wafer fabrication the wafers pass many times in and out of four basic operations (layering, patterning, doping, and heat processing). In packaging also there are several basic operations (Fig. 18.8). However, packaging is a once-through process; each of the major processes is required only once. As in fabrication, the exact order of the operations is determined by the package type and other factors. Each operation may or may not be used in a particular process and each is customized to the particular chip and package requirements.

The basic operations are *backside preparation*, *die separation*, *die pick*, *inspection*, *die attach*, *wire bonding*, *preseal inspection*, *package*

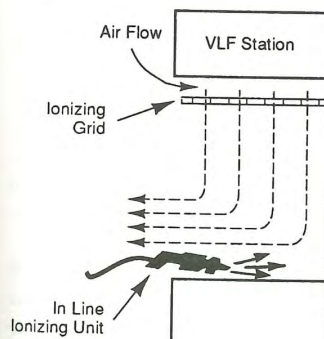


Figure 18.7 Static control techniques.

- | | |
|------------------------|----------------------|
| • Backside Preparation | • Preseal Inspection |
| • Die Separation | • Packaging Sealing |
| • Die Pick | • Plating |
| • Die Attach | • Lead Trim |
| • Inspection | • Marking |
| • Bonding | • Final Tests |

Figure 18.8 Basic packaging operations.

sealing, plating, trimming, marking, and final tests. The details of the various operations and their optional processes are explained in more detail in the remainder of this chapter.

Backside preparation. At the end of the fabrication process some wafers have to be thinned (wafer thinning) to fit in the package and/or to remove backside damage or junctions. Wafers whose die are going to be attached to the package by a gold-silicon eutectic will receive a deposited layer of gold (backside gold).

Die separation. The wafer is separated into individual die by sawing or scribe-and-break techniques.

Die pick and plate. The functioning die identified at wafer sort are picked from the separated wafer and placed in carriers (Fig. 18.9).

Inspection. The plated dies are optically inspected for edge integrity, contamination, and cosmetic defects.

Die attach. Each die is attached into the die-attach area of the package by either a gold-silicon eutectic layer or an epoxy adhesive material (Fig. 18.10).

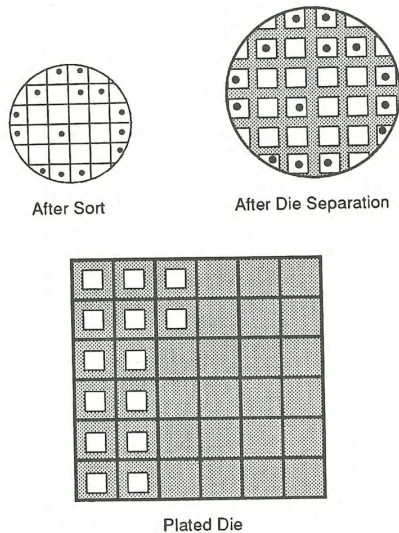


Figure 18.9 Die separation to plate.



Lead bo
and the
metal o

Preseal
age. Cri
proper k
ing qual

Plating,
with an
ability in
package
arate the
code imp
enclosure

Final test.
tal, are p
packages



Figure 18.11

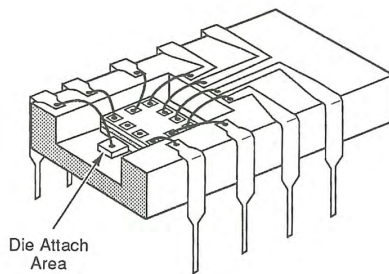


Figure 18.10 Die attach area of package.

Lead bonding. Thin wires are bonded between the chip bonding pads and the inner leads of the package. Bonding is also done by bumps of metal on the bonding pads or by the TAB technique.

Preseal inspection. The bonded die is optically inspected in the package. Criteria checked for are the alignment of the die in the package, proper bond placement, contamination, die-attach quality, and bonding quality.

Plating, trimming, and marking. The outer package leads are plated with an additional layer of conductive metal to improve their solderability into a printed-circuit board. Near the end of the process, each package with outer leads will go through a trimming operation to separate the leads from supports. The marking operation is performed to code important information (Fig. 18.11) on the outside of the package enclosure.

Final test. A series of final tests, including electrical and environmental, are performed on each packaged chip to ensure quality. Some packages will receive a "burn-in" test to detect early failures.



Figure 18.11 Marked package.

Common package parts

The four functions of a chip package are accomplished through the use of a wide variety of package designs. However, most packages have five common parts: the die-attachment area, the inner leads, the outer leads, the chip to package connection, and the enclosure.

Die-attachment area. In the center of every package is an area where the chip is securely attached into the package. This die-attachment area may have an electrical connection that serves to connect the back of the chip to the rest of the lead system. A major requirement for this area is absolute flatness in order to intimately support the chip in the package (Fig. 18.10).

Inner and outer leads. The metal lead system chip is continuous from the die-attach cavity to the printed-circuit board or electronic product. The system inner connections are called the *inner leads*, *bonding lead tips*, or *bond fingers*. The inner leads are generally the narrowest portion of the lead system. The leads become progressively wider or thicker, finally ending outside the package. These portions of the lead system are called the *outer leads* (Fig. 18.12). Most of the lead systems are composed of one continuous piece of metal. One exception is the side-brazed package. In this package construction method, the outer leads are brazed onto the interior leads. Two different alloys are used for the outer lead system, either an iron-nickel alloy or a copper alloy. The iron-nickel alloy is desirable for its strength and stability, while copper is used for its electrical and heat-conduction properties.

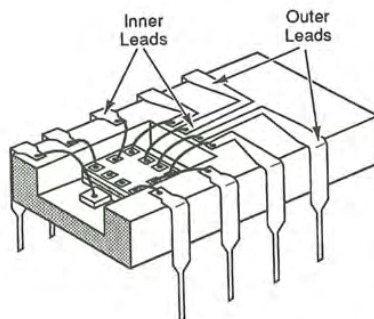


Figure 18.12 Inner and outer leads.

Bonding W



Chip to package

The chip is attached to the package with bonding wires.

Enclosure

The enclosure is the electronic package body. This enclosure has various functions: protection, hermetic sealing, and heat dissipation.

Hermetic. Hermetic sealing is a technique used to completely seal the package from the environment, preventing moisture and contaminants from entering.

Nonhermetic. Nonhermetic packages are not sealed and are used in applications where hermetic sealing is not required.

- | |
|-------------|
| Hermetic |
| • Metal |
| • Ceramic |
| Nonhermetic |
| • Epoxy |
| • Polyimide |

Figure 18.14

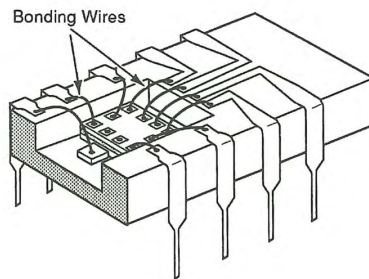


Figure 18.13 Bonding wires.

Chip to package connection

The chip is electrically connected to the lead system of the package with bonding wires, balls, or other on-chip connectors (Fig. 18.13).

Enclosures. The die-attach area and inner and outer leads constitute the electrical parts of the package. The other part is the *enclosure* or *body*. This is the part that provides the protection and heat-dissipation functions. These functions are achieved by several different techniques and package designs as described in the sealing section. The completeness of the seal falls into two categories, hermetic and nonhermetic (Fig. 18.14).

Hermetic sealing results in a package that is impervious to the penetration of moisture and other gases. Hermetic seals are required for chips operating in harsh and demanding environments such as rockets or space satellites. Metal and ceramic enclosures are the preferred materials for hermetically sealed packages.

Nonhermetically sealed packages are adequate for most consumer applications such as computers and entertainment systems. This sealing system provides good and adequate environmental protection of the chip, except in the most demanding situations. A better term

Hermetic
• Metal
• Ceramic
Nonhermetic
• Epoxy Resins
• Polyimides

Figure 18.14 Package sealing designations.

for this type of enclosure sealing method would be "less hermetic." Nonhermetic packages are composed of epoxy resins or polyimide materials and are generally referred to as "plastic packages."

Packaging Processes

Prepackaging wafer preparation

After the final passivation layer and an alloy step in wafer fabrication, the circuits are complete. However, one or two additional processes may be performed on the wafer before transfer to packaging. These steps, wafer thinning and backside gold, are optional, depending on the wafer thickness and the particular circuit design.

Wafer thinning. The trend to thicker wafers presents several problems in the packaging process. Thicker wafers require the more expensive complete saw-through method at die separation. While sawing produces a higher-quality die edge, the process is more expensive in time and consumption of diamond-tipped saws. Thicker die also require deeper die attach cavities, resulting in a more expensive package. Both of these undesirable results are avoided by thinning the wafers before die separation. Another situation requiring wafer thinning is electrical in nature. If the wafer backs are not protected as the wafers go through the dopant operations in fabrication, the dopants will form electrical junctions in the wafer back, which may interfere with good conduction in the back contact, which is required for the circuit to operate correctly. These junctions may require physical removal by wafer thinning.

The thinning step generally takes place between wafer sort and die separation. Wafers are reduced to a thickness of 0.2-0.5 mm.¹ Thinning is done by the same processes (mechanical grinding and chemical-mechanical polishing or CMP) used to grind wafers in the wafer preparation stage (see Chap. 3). An alternate method is to protect the front of the wafers and chemically etch material from the back.

Wafer thinning is a worrisome process. In backgrinding there is the concern of scratching the front of the wafer and of wafer breakage. Since the wafer must be held down on the grinder or polishing surface, the front of the wafer must be protected, and once thinned, wafers break more easily. In backetching there is a similar need to protect the front of the wafer from the etchant. The protection can be provided by spinning a thick layer of photoresist on the front side. Other methods include attaching adhesive-backed polymer sheets cut

to the wafer and polish and warping. (broken and the package

Backside gold. The backside gold can be attached to the "attach" section (after back

Die separation

The chip is separated into individual dies and sawing

Scribing. Scribe separation is a method of the wafer separation. The diamond-tipped scribe creates a shallow groove in the die is completed. The roller is used to create scribe lines. creating a ridge on the wafers over

Sawing. The sawing as the wafer is moved across the wafer table. The system for the round saw. The

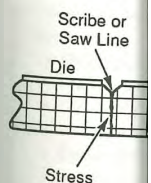


Figure 18.15 Scribe

to the wafer diameter. Stresses induced in the wafer by the grinding and polishing processes must be controlled to prevent wafer and die warping. Wafer warping interferes with the die separation process (broken and cracked die). Die warping creates die-attach problems in the packaging process.²

Backside gold. Another optional wafer process is adding a layer of backside gold. A layer of gold is required on wafers that are going to be attached to the package by eutectic techniques (see the "Die attach" section). The gold is usually applied in the fabrication area (after backgrinding) by evaporation or sputtering.

Die separation

The chip-packaging process starts with the separation of the wafer into individual dies. The two methods of die separation are scribing and sawing (Fig. 18.15).

Scribing. *Scribing, or diamond scribing,* was the first production die-separation technique developed in the industry. It requires the alignment of the wafer on a precision stage, followed by dragging a diamond-tipped scribe through the center of the scribe lines. The scribe creates a shallow scratch in the wafer surface. The separation of the die is completed by stressing the wafer with a cylindrical roller. As the roller is moved over the surface, the wafer separates along the scribe lines. The breaks follow the crystal structure of the wafer, thus creating a right-angle edge on the die. Scribing becomes unreliable in wafers over 10 mils thick.

Sawing. The advent of thicker wafers has led to the development of sawing as the preferred die-separation method. A saw consists of a wafer table with rotation capability, a manual or automatic vision system for orienting the scribe lines, and a diamond-impregnated round saw. Two techniques are used. Both start with the passage of

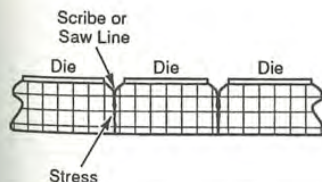


Figure 18.15 Scribe and saw separation.

the diamond saw over the scribe lines. For thinner wafers, the saw is lowered into the wafer surface to create a trench about one-third of the way through the wafer. The separation of the wafer into die is completed by the stress and roller technique used in the scribing method. The second sawing method is to separate the die by a complete saw-through of the wafer.

Often the wafers undergoing complete saw-through are first mounted on a flexible plastic film. The film holds the die in place after the sawing operation and aids the die-pick operation. Sawing is the preferred die-separation method due to the cleaner die edges and the fewer cracks and chips left on the sides of the die (Fig. 18.16).

Die pick and plate

After sawing, the separated die are transferred to a station for selection of the functioning die (nonlinked). In the manual method, an operator will pick up each of the nonlinked dies with a vacuum wand and place it in a sectioned plate. Wafers that come to the station on the flexible film are first placed on a frame that stretches the film. The stretching separates the die, which aids the die-pick operation.

In the automated version of this operation, a memory tape or disk with the locations of the good die (from wafer sort) is loaded into the tool. A vacuum wand picks up good die and automatically places them in a sectioned plate for transfer to the next operation.

Die inspection

Before being committed to the rest of the process, the die are given an optical inspection. Of primary interest is the quality of the die edge, which should be free of chips and cracks. This inspection also sorts out surface irregularities, such as scratches and contamination. Inspection may be manual with microscopes or automated with a vision system. At this step the die are ready to go into a package.

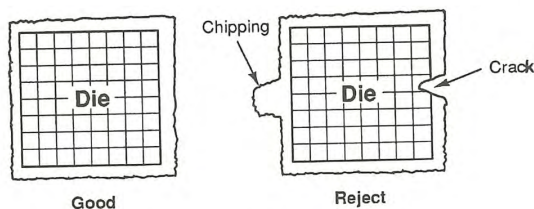


Figure 18.16 Die-separation results.

Die att

Die at
betwe
ductin
serve a

A re
bond s
steps
require
ed to h
the die
outgas
itself s

Ther
adhesi

Eutectic
non the
much l
attach,
Gold m
are mix
onto the
when he

The g
of the c
Someti
mixture
ers, alon
thin allo
attachm
Eutect
package

Conducti
• Gold
• Meta
• Conc
Noncond
• Epox
• Insula

Die attach

Die attachment has several goals: to create a strong physical bond between the chip and the package, to provide either an electrical conducting or insulating contact between the die and the package, and to serve as a medium to transfer heat from the chip to the package.

A requirement is the permanency of the die-attachment bond. The bond should not loosen or deteriorate during subsequent processing steps or when the package is in use in an electronic product. This requirement is especially important for packages that will be subjected to high physical forces such as those used in rockets. Additionally, the die-attach materials should be contaminant-free and should not outgas during subsequent process heating steps. Lastly, the process itself should be productive and economical.

There are two principal methods of die attach, *eutectic* and *epoxy adhesives*.

Eutectic die attach. The eutectic method is named for the phenomenon that takes place when two materials melt together (alloy) at a much lower temperature than either of them separately. For die attach, the two eutectic materials are gold and silicon (Fig. 18.17). Gold melts at 1063°C, while silicon melts at 1415°C. When the two are mixed together, they start alloying at about 380°C. Gold is plated onto the die-attach area and alloys with the bottom of the silicon die when heated.

The gold for the die-attach layer is actually a sandwich. The bottom of the die-attach area is deposited or plated with a layer of gold. Sometimes a preformed piece of metal composed of a gold and silicon mixture is placed in the die-attach area. When heated, these two layers, along with a thin layer of silicon from the wafer back, forms a thin alloy layer. This layer is the actual bond forming the die-package attachment.

Eutectic die attach requires four actions. First is the heating of the package until the gold-silicon forms a liquid. Second is the placement

- | |
|---|
| <p>Conducting</p> <ul style="list-style-type: none"> • Gold/Silicon Eutectic • Metal-Filled Epoxy • Conducting Polyimide <p>Nonconducting</p> <ul style="list-style-type: none"> • Epoxy Adhesive • Insulating Polyimide |
|---|

Figure 18.17 Die-attach material matrix.

Epoxy die attach. The alternate die-attach process uses thick liquid epoxy adhesives. These adhesives can form an insulating barrier between the chip and package or become electrically and heat conducting with the addition of metals such as gold or silver. Polyimide may also be used as an adhesive. Popular adhesives are silver-filled epoxy for copper lead frames and silver-filled polyimide for alloy 22 metal frames.³

Epoxy die attach is favored for its economy and ease of processing, in that the package does not have to be heated on a stage. This factor makes the automation of the process easier. When compared to gold-silicon eutectic die attach, epoxy has the disadvantage of potential decomposition at the high temperatures of bonding and sealing operations. Epoxy die-attach films also do not have the bonding power of the eutectics.

Bonding

Once the
ing pro
operati
connect

- Wire
- Flip-
- TAB

The v
1.0-mil)
the inne
the wire
process
and pro
placeme
rate pla
tact at l
scribed l
wires. W
those in
adjacent

Wire b
highly co
ing the b
its advan
methods.

Gold wire
al. It is t
lent heat
translated
aluminum
methods a
thermoson

Thermocouples
with the
heating coils
that are
processed
frame only

Bonding pad and package connection

Once the die and package are attached, they go to the bonding process. This is perhaps the most critical of all the assembly operations. Three techniques provide the critical chip to package connection:

- Wire bonding
- Flip-chip and bump
- TAB

The wire bonding procedure is simple in concept. A thin (0.7- to 1.0-mil) wire is first bonded to the chip bonding pad and spanned to the inner lead of the package lead frame. The third action is to bond the wire to the inner lead. Last, the wire is clipped and the entire process repeated at the next bonding pad. While simple in concept and procedure, wire bonding is critical because of the precise wire placement and electrical contact requirements. In addition to accurate placement, each and every wire must make good electrical contact at both ends, span between the pad and inner lead in a prescribed loop without kinks and be at a safe distance from neighboring wires. Wire loops in conventional packages are 8 to 12 mils, while those in ultrathin packages are 4 to 5 mils.⁴ The distance between adjacent wires is called the pitch of the bonding.

Wire bonding is done with either gold or aluminum wires. Both are highly conductive and ductile enough to withstand deformation during the bonding steps and still remain strong and reliable. Each has its advantages and disadvantages and each is bonded by different methods.

Gold wire bonding. Gold has several pluses as a bonding wire material. It is the best-known room-temperature conductor and is an excellent heat conductor. It is resistant to oxidation and corrosion, which translates into an ability to be melted to form a strong bond with the aluminum bonding pads without oxidizing during the process. Two methods are used for gold bonding. They are *thermocompression* and *thermosonic*.

Thermocompression bonding (also known as *TC bonding*) starts with the positioning of the package on the bonding chuck and the heating of the chip and package to between 300 and 350°C. Chips that are going to be enclosed in an epoxy molded package are processed through die attach and bonded with the chip on the lead frame only. The bonding wire is fed out of a thin tube called a *capil-*