

SYSTEM LEVEL

LOGIC LEVEL

CIRCUIT LEVEL

PHYSICAL LEVEL

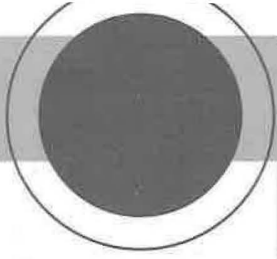
FABRICATION

Chip Design for Submicron VLSI:

CMOS Layout and Simulation

John P. Uyemura

W/ 1CD



Chip Design for Submicron VLSI: CMOS Layout and Simulation

John P. Uyemura

Georgia Institute of Technology

THOMSON



TM

Australia Canada Mexico Singapore Spain United Kingdom United States

THOMSON



Chip Design for Submicron VLSI: CMOS Layout and Simulation

by John P. Uyemura

**Associate Vice-President and
Editorial Director:**

Evelyn Veitch

Publisher:

Bill Stenquist

Sales and Marketing Manager:

John More

Developmental Editor:

Kamilah Reid Burrell

Production Services:

RPK Editorial Services

Copy Editor:

Shelly Gerger-Knecht

Proofreader:

Pat Daly

Indexer:

Melba Uyemura/RPK Editorial
Services

Production Manager:

Renate McCloy

Creative Director:

Angela Cluer

Interior Design:

Carmela Pereira

Cover Design:

Katherine Strain

Compositor:

PreTeX, Inc.

Printer:

Transcontinental Printing

Cover Photo:

Yalcin Alper Eken,
Atlanta, Georgia

COPYRIGHT © 2006 by Nelson, a division
of Thomson Canada Limited.

Printed and bound in Canada

1 2 3 4 07 06 05 04

For more information contact Nelson,
1120 Birchmount Road, Toronto, Ontario,
Canada, M1K 5G4.

Or you can visit our Internet site at
<http://www.nelson.com>

Library of Congress Control Number:
2004117356

ISBN: 0-534-46629-X

ALL RIGHTS RESERVED. No part of this
work covered by the copyright herein may
be reproduced, transcribed, or used in any
form or by any means—graphic, electronic,
or mechanical, including photocopying,
recording, taping, Web distribution, or infor-
mation storage and retrieval systems—with-
out the written permission of the publisher.

For permission to use material from this text
or product, submit a request online at
www.thomsonrights.com.

Every effort has been made to trace owner-
ship of all copyright material and to secure
permission from copyright holders. In the
event of any question arising as to the use of
any material, we will be pleased to make the
necessary corrections in future printings.

North America

Nelson
1120 Birchmount Road
Toronto, Ontario M1K 5G4
Canada

Asia

Thomson Learning
5 Shenton Way #01-01
UIC Building
Singapore 068808

Australia/New Zealand

Thomson Learning
102 Dodds Street
Southbank, Victoria
Australia 3006

Europe/Middle East/Africa

Thomson Learning
High Holborn House
50/51 Bedford Row
London WC1R 4LR
United Kingdom

Latin America

Thomson Learning
Seneca, 53
Colonia Polanco
11560 Mexico D.F.
Mexico

Spain

Paraninfo
Calle/Magallanes, 25

Views of a Chip—Layers and Patterns

Chip design is a fascinating field to study and work in. Using a sophisticated set of computer-aided design (CAD) tools, one can envision a universe that is completely different from the one we live in. In this microscopic world, electrons travel through reconfigurable mazes that perform logic and are created by the chip designer. In this chapter, we will introduce you to the world of a chip designer as seen at the physical silicon level.

2.1 The Design Hierarchy

Integrated circuits (ICs) are microscopic electronic networks that are created in a special type of material called a **semiconductor**. Silicon is a semiconductor and is used as the base material for the vast majority of modern electronic systems. As integrated circuits became more commonplace, writers started using the phrase **computer chip** to describe the devices used to create the computer revolution, and the term has evolved to be accepted by all. The descriptive word *chip* arises from the physical appearance of the prototype silicon integrated circuit with both digital and analog circuit functions. Figure 2.1 is a photograph of a silicon integrated circuit that has side dimensions of 2.7 mm × 2.6 mm.¹ The actual size of the circuit is shown here:

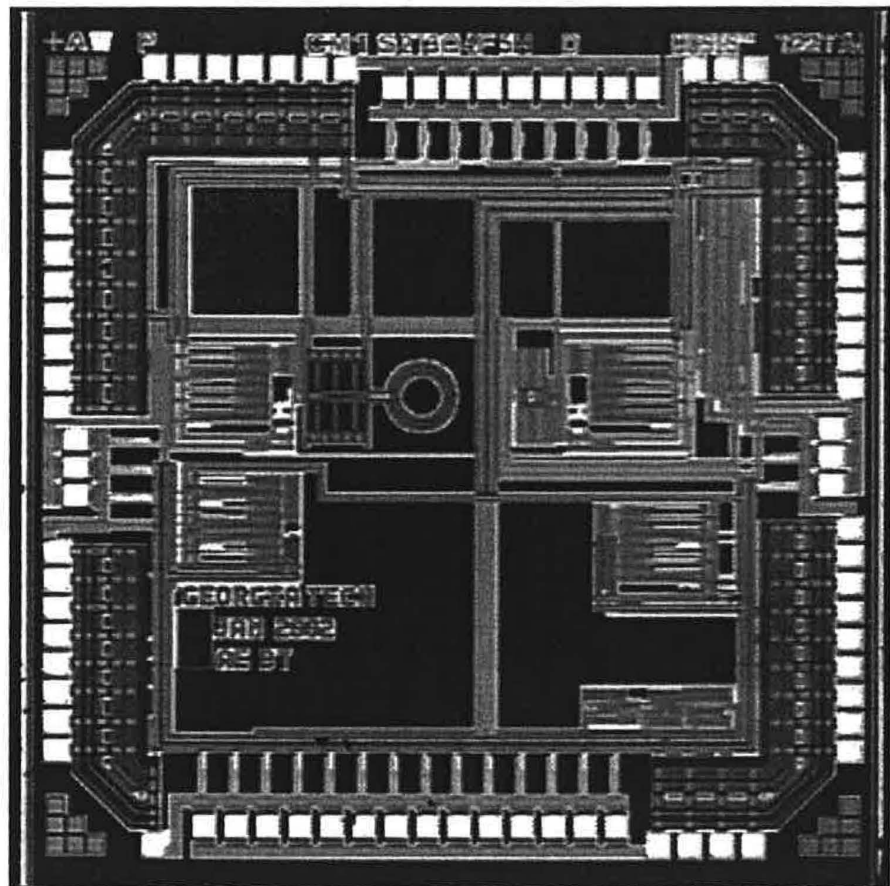


This example shows that it really does look like a “chip” of some sort!

Although you may think that computer chips have been around forever, integrated circuits were not invented until 1956 by Jack Kilby of Texas Instruments Corporation. The

1. This chip was designed in a joint IDT-Georgia Tech project by Dr. Brian Butka (IDT), Dr. Y. Alper Eken, Dr. Paul Murtagh (IDT), and Dr. Bortecene Terlemez.

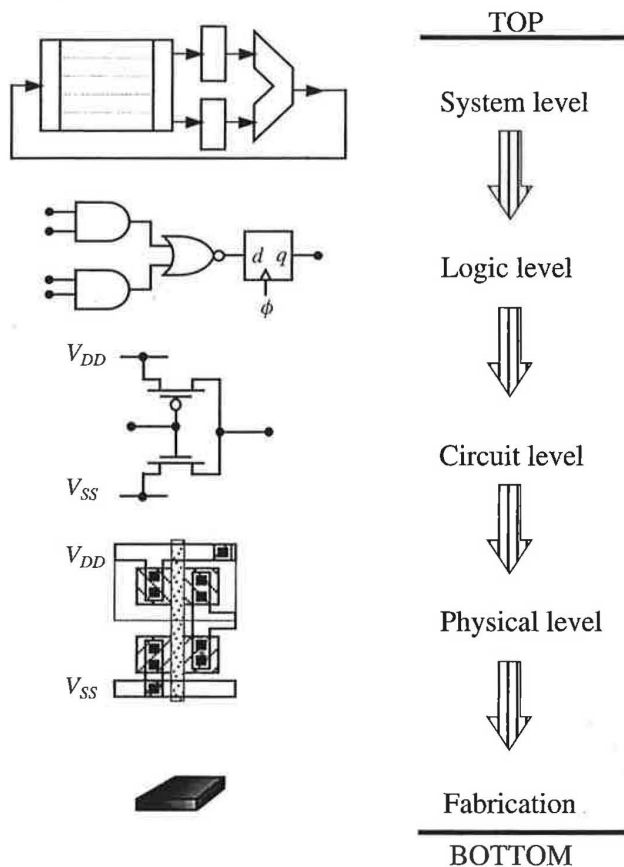
Figure 2.1:
Photograph of an
experimental CMOS
telecommunication
chip



significance of this invention is reflected by the fact that Kilby shared the 2000 Nobel Prize in physics for his work. Integrated circuits gained steady acceptance and usage as our knowledge and design skills increased and the cost per chip decreased. The development of the personal computer (PC) in the 1980s fueled the “computer revolution,” and the unparalleled explosive growth of modern computer systems and networks has provided the driving force to the present day.

Integrated circuits are quite complex, and a complete understanding of every aspect of chip design and fabrication requires several years of study and practical experience. This task is made easier by breaking the problem into **design hierarchies** where the problem is viewed at several different levels. Each level concentrates on a specific viewpoint, such as logic design, that can be linked to other levels. Some levels are abstract, while others can

Figure 2.2:
Structure of the
design hierarchy



be visualized very easily. Figure 2.2 shows the basic breakdown of the hierarchy for a digital integrated circuit. Starting at the top of the drawing, we see the highest level is:

- **System Design**—Where the main operations of the chip are determined. Block diagrams are used to illustrate the main sections that make up the system. Only the input/output characteristics are important, and there are no details about what actually is inside each block.

The next level down is that of

- **Logic Design**—This is where the engineering teams design the logic networks that are required inside each block to obtain the input/output characteristics used at the system design level. The output is generally in the form of a netlist, which is just a description of the logic gates and wiring needed to implement the design.

Once this is accomplished, we move to the next lower level of

- **Circuit Design**—Where the logic network is transformed into an electronic network using transistors as switching devices. Digital variables are represented by voltage levels that change in time. Transistors allow the designer to create logic circuits that steer signals into different paths using switching mechanisms.

Once the circuit design is completed, it is transformed into a design file for a silicon integrated circuit in the next level that is designated as

- **Physical Design**—In this step, the electronic circuits are transformed into on-screen colored geometrical patterns using computer graphics and analysis tools. Each color or shading represents a material, such as a metal, and the patterns indicate how to form 3-dimensional transistor structures and wire them together.

The final step is that of

- **Chip Fabrication**—Where the physical design is transformed into a finished silicon chip that can be put into a package for eventual wiring into a product.

This sequence illustrates the **top-down** approach to chip design that is used for modern digital-chip designs. Project engineering is performed by groups at every level, and information exchanged among the various design levels is crucial. Many sub-tasks will overlap, but the hierarchy usually remains visible. The size of the project group varies with the complexity of the chip. While a basic, functional chip may be completed using only four or five main designers, a state-of-the-art microprocessor chip can require several hundred engineers, technicians, and support personnel for successful completion.

This book does not attempt to cover the entire design hierarchy, but concentrates on the study of physical design as a central theme. However, the close-knit characteristic of the design hierarchy leads us to other levels in a natural manner. This is because physical design is based on the characteristics of the fabrication sequence and forms the link to circuit and logic design. The Microwind software provides this link in an integrated manner and allows us to see the relationship among these three levels in a seamless manner.

2.2 Integrated Circuit Layers

In its simplest form, an integrated circuit can be viewed as a set of *patterned material layers*. If you worked through the examples in Chapter 1, then this statement already has some meaning. It does, however, deserve a more detailed discussion before we go into the intricacies of layout design.

A chip is the physical implementation of a signal control or processing unit. The signals change in time and are represented by voltages $V(t)$ at different points on the network.

Figure 2.3:



An example is shown in Figure 2.3. If we assume that the voltage can vary from 0 V to a peak value of V_{DD} , then we can associate binary equivalents with different ranges. In CMOS, the standard designation for the power supply voltages applied to a chip are

V_{DD} = positive power supply voltage

V_{SS} = ground or 0 V

The actual value of V_{DD} varies with the fabrication process. Some common values are $V_{DD} = 5$ V, 3.3 V, 2.5 V, and 1.2 V, but others are also found in practice.

The **positive logic** convention assigns a binary "0" to the low voltages, while a binary "1" is represented by high voltages. The intermediate values have no meaning in the Boolean world. Theoretical considerations provide the large-scale model of the system, but the final realization is a network that operates on the basic principles of currents and voltages.

To a chip designer, every material is characterized by how well it conducts electrical current. Metals are classified as **conductors**, meaning that they conduct electricity very well. Aluminum (Al) and copper (Cu) are the most common conductors used in CMOS integrated circuits. An **insulator** is exactly the opposite: it resists the flow of electrical current. If you examine a section of "wire," such as the type that is used to connect speakers to an amplifier, you will see that the electrical current flows on a metal strand that is encased by a plastic insulating jacket. Figure 2.4 shows this structure. The insulating cover insures that unwanted conduction paths will not be set up between two lines that happen to touch.

Figure 2.4:
Structure of a
speaker wire

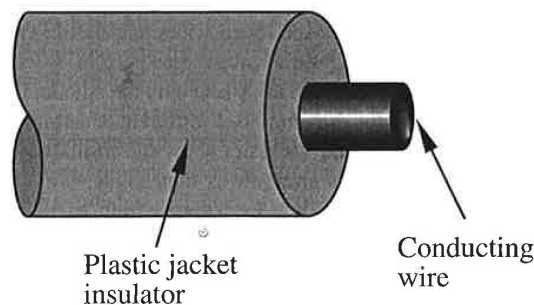
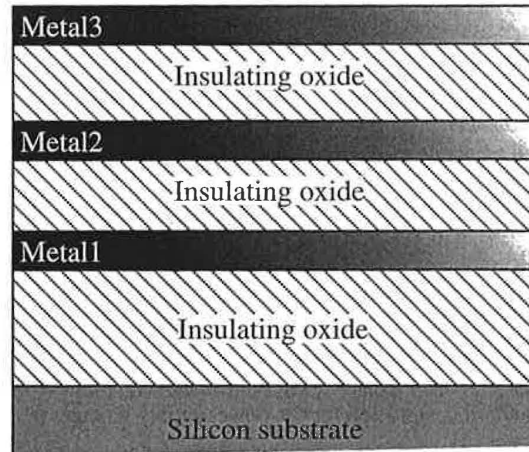


Figure 2.5:
Metal and insulator
layers of a chip

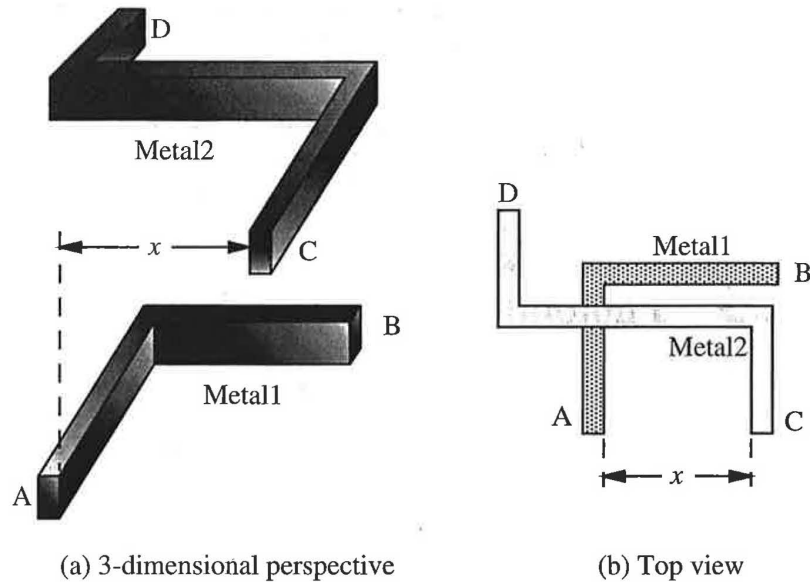


This simple idea provides the basis for integrated circuit design. Starting with a flat-surfaced area of silicon, we use various techniques to create an alternating stack of thin metal and insulating layers, as portrayed in Figure 2.5. The silicon itself provides the base for the structure, and is called the **substrate**; it is usually about 0.3–0.5 mm thick. The substrate is coated with a layer of insulating material such as **silicon dioxide (SiO_2)**, otherwise known as quartz glass. In our drawings, the insulating layers are generally referred to as oxides or glass, regardless of their actual composition.

The metal layer closest to the silicon is called Metal1, since it is the first to be deposited on the structure. After another layer of oxide is added, the second layer of metal (Metal2) is deposited on top. This is followed by another layer of insulating oxide, which is followed by another layer of metal shown as Metal3 in the drawing. This alternate oxide–metal layering scheme can be continued, and modern chips have eight or more layers of conducting metal.

Each metal layer is electrically isolated from all others by the insulating oxides. If we go one step further and define distinct patterns of **conduction lines** on the individual metal layers, then we can independently steer electrical signals on each level. This is analogous to creating small rectangular cross-sectioned wires, and is illustrated in Figure 2.6. The view in Figure 2.6(a) shows a 3-dimensional perspective for distinct patterns on the Metal1 and Metal2 layers. When these are stacked on the silicon, they give us the top view shown in Figure 2.6(b). This drawing uses different shadings/patterns to distinguish between the two metal layers, and we don't show the insulating oxide (since it is glass anyway). The significance of this structure is that points A and B are electrically connected through the Metal1 pattern, but are isolated from Metal2 points C and D. If we want a Metal1 line to contact a Metal2 line, then we can create a hole in the insulating oxide layer and fill it with metal that touches both metal layers. This is generally called a **contact** or a **via** and allows us to send signals on any desired layer.

Figure 2.6:
Patterned metal
layers in an
integrated circuit



A silicon integrated circuit uses patterned metal lines to provide conduction paths between points in the electronic circuit. Having access to several metal layers eases the problem of connecting a complex circuit together. Switching functions are implemented using controlled switching devices called **transistors** that are built into the silicon level by creating patterns of regions with different conduction characteristics. Control is achieved using a conducting layer above the silicon level. As you can see, a silicon chip really is just a set of patterned material layers.

We are finally in a position to define what this book is about. A digital integrated circuit is a logical switching network that consists of transistors wired together in a very specific way. The physical design of a silicon chip centers on creating patterns on each layer of an integrated circuit such that the resulting 3-dimensional structures form transistors and the wiring that connects them together. The shapes and sizes of each pattern are important to determining the electrical performance of the chip. This includes parameters such as the switching times, the power consumption, and the total area of the chip. Launching Microwind places you in a **layout editor**. This is a CAD² tool that is used to design the patterned layers of the integrated circuit. It gives you the ability to do physical design in a graphical environment. The electrical characteristics are obtained by using a circuit simulator that is built into the Microwind package. Our treatment will teach you the fundamentals of physical design as it is constrained by the fabrication process. We will also examine how our design translates into electrical characteristics using a SPICE³ circuit simulator.

2. CAD is an acronym for Computer-Aided Design.

3. SPICE is an acronym for Simulation Program with Integrated Circuit Emphasis.

The treatment will include both text discussions and hands-on PC examples that have been designed to reinforce your understanding of the material. Physical design is best learned by hands-on experience, so you should make an effort to follow the examples on your computer. Microwind is unique in that it integrates many useful engineering and pedagogical features in a seamless manner. You will probably find it extremely useful for studying integrated electronics far beyond what is covered here.

2.3 Photolithography and Pattern Transfer

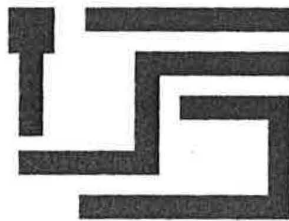
A key concept in chip design revolves around the ability to create patterned material layers to steer electrical signals on the chip. This is accomplished using a technique called **photolithography** that allows us to translate on-screen computer drawings to a physical structure that replicates the patterns defined by our CAD package. The resolution of the photolithographic process determines the minimum feature size that can be fabricated on the chip, and is a limiting factor in the design of high-density integrated circuits. The sequence that is used to create the pattern can be summarized as

- Design the pattern on a computer
- Create a mask or reticle
- Print the pattern onto the surface of the chip
- Use the printed region to define the material pattern

Since this may seem a bit mysterious, let us delve deeper into this important aspect of chip design and manufacturing. Constraints introduced by photolithography have a great effect on how we design circuit layouts, making this a priority in our study.

The starting point is the computer data base that defines each polygon and its location on a layer. This information is translated into a standardized format that identifies each geometrical object by its vertices. Figure 2.7 illustrates the conversion to a Caltech intermediate form (CIF) file format. The file is then used to create a **mask** or **reticle** that is used to optically transfer the original pattern to the chip surface. The structure of a basic

Figure 2.7:
Translation of a
pattern to a CIF file

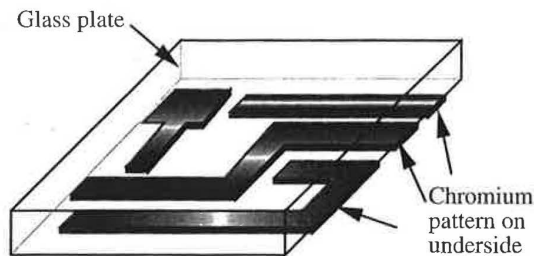


(a) On-screen plot

```
L 13;
P 780,3360 960,3360 960,4020 780,4020;
P 10380,-420 10560,-420 10560,840 10380,840;
P 9600,-780 9960,-780 9960,-420 9600,-420;
P 2640,2040 2760,2040 2760,3360 2640,3360;
P 10380,3360 10560,3360 10560,4020 10380,4020;
P 780,1560 960,1560 960,2040 780,2040;
P 9600,4020 9960,4020 9960,4380 9600,4380;
```

(b) CIF file description

Figure 2.8:
Structure of a
basic reticle

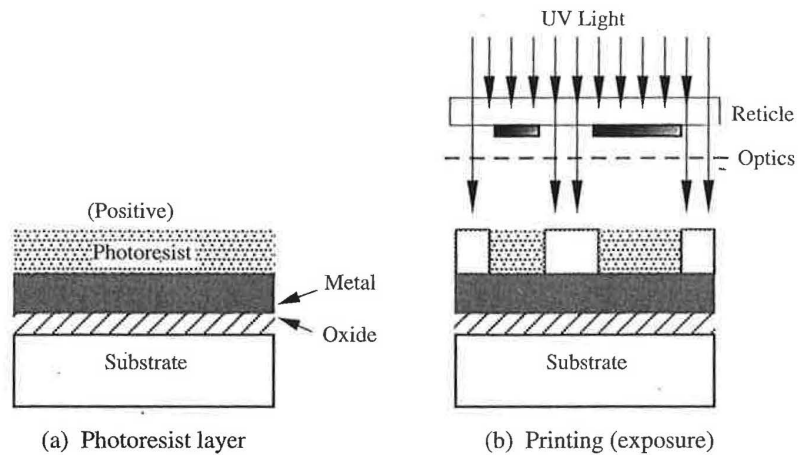


reticle is shown in Figure 2.8. It consists of a high-quality piece of glass with a chromium metal replica of the pattern on one side. The dimensions of the reticle pattern are larger than the final values that will be transferred to the chip. This allows finer precision for small-geometry devices. For use in photolithography, the main idea is that the reticle is optically transparent except in those regions where the chromium metal regions exist; there, the reticle is opaque and any incident light is reflected by the metal. In the chip manufacturing process, each layer is patterned separately from every other layer. Thus, one mask/reticle is required for each layer. Mask sets for an entire chip are extremely expensive to fabricate, with costs exceeding \$1,000,000 per set not uncommon in a state-of-the-art design. One reason for the high cost is that the features on a modern process are so small, a simple transmission/reflection technique is not sufficient to define the details. Many advanced optical methods based on diffraction and other effects must be used to create a good reticle. This complicates the process considerably, but it is a mandatory part of the chip-making procedure.

Transferring the reticle pattern to the surface of an integrated circuit takes place in the exposure sequence. The process starts by coating the surface of the chip with a light-sensitive organic polymer (i.e., a plastic) called **photoresist** (or just **resist**) which acts similarly to ordinary photographic film. When illuminated by an ultraviolet (UV) light source, regions of the photoresist that absorb optical energy experience an internal chemical change; if we can shield regions from the light source, then they are not affected. Developing the photoresist in a special solution yields a layer where some regions are hardened, while other regions can be washed away with a solvent. There are two "types" of photoresist, positive and negative. In positive resists, the illuminated regions are soluble after exposure to light, while unexposed regions develop into hardened material. Negative resists are exactly the opposite: the regions that are exposed to light develop into hardened sections. The clear and opaque regions of the reticle must be specified in the layout and transferred to the reticle in the mask-making process.

Consider the sequence described by the cross-sectional views in Figure 2.9. The starting point in this sequence is a substrate with an oxide coating and a metal layer on top of the oxide. Our objective is to transfer a pattern to the metal layer. First, photoresist is applied by spin-coating the surface of the chip with a liquid solution, and then baking it for a few minutes to dehydrate the layer. This leaves a photoresist coating, as in Figure 2.9(a). Next, the reticle is placed between a UV light source and the chip so that it casts a shadow onto the surface of the photoresist; Figure 2.9(b) shows this step. Optical imaging components (not shown) are used to sharpen the edges of the shadow to insure a high-quality

Figure 2.9:
Resist application
and exposure



print. This is called the **exposure step**, and only requires a few seconds to complete. After the resist is exposed, it is developed and rinsed in a solvent; this leaves a hardened layer of photoresist that has the same pattern as the original screen image.

The final structure after the developing procedure is shown in Figure 2.10. The important aspect of the drawing is the fact that the resist layer now contains the pattern information that was designed into the layout. Hardened photoresist itself is used as a masking layer in the final step of the lithographic-transfer process. It acts as an efficient barrier to chemical or gaseous agents that can be used to etch away the metal regions. This is called the **etching process** and is shown in Figure 2.11. Modern IC fabrication employs **Reactive Ion Etching (RIE)** in which ionized atoms of an inert gas such as argon (Ar) are mixed with etch-assisting chemicals. The mixture is then excited with a radio frequency (rf) electric field in a manner that drives the ions/chemicals in a vertical up-down motion to etch away the surface; this is portrayed in Figure 2.11(a). Unprotected metal surfaces are removed by the process, but the resists protects any metal underneath. This results in the cross-section shown in Figure 2.11(b). Removing the photoresist gives the final result: a patterned metal layer that has the same features as those designed on the computer. This is shown in Figure 2.11(c).

Figure 2.10:
After development

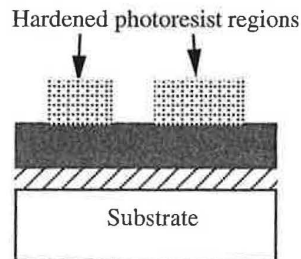
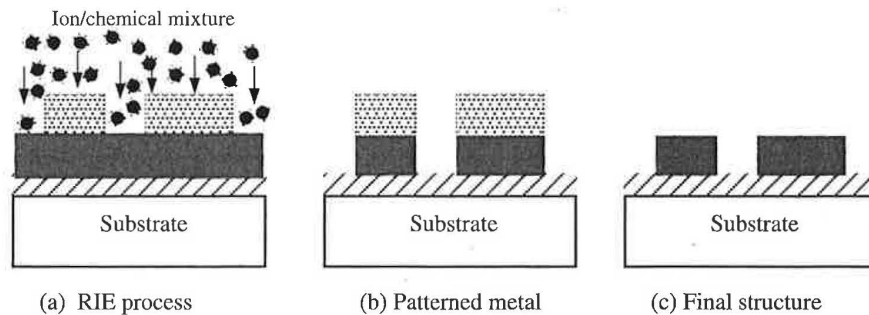
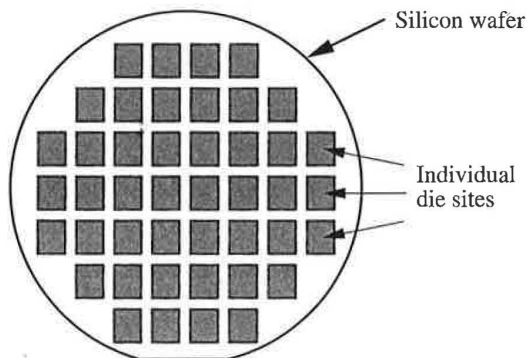


Figure 2.11:
Etching process

All material layers above the substrate are patterned using this sequence. This includes all metals and oxides (for holes). We will also need to modify some regions on the silicon substrate to make the transistor switches. To accomplish this, we use the same procedure to transfer the pattern, but change the processing steps to create the desired structures. This will be discussed later when we examine the entire process flow for an integrated circuit.

Now that we understand how the printing process works, let us take a short trip to the world of chip manufacturing to see how the process is modified for mass production. Integrated circuits are fabricated on large silicon wafers that are 200–300 mm (8–12 inches) in diameter (or larger), and about 1 mm thick. Many ICs are simultaneously created on a single wafer, as shown in Figure 2.12. The number of individual devices (or *die*) depends upon the diameter of the wafer and the shape and area of each die site. Obviously, a larger diameter wafer allows us to manufacture more ICs per wafer, but increasing the diameter makes uniform processing more difficult to achieve. When a reticle is used to pattern a layer, each die site is treated separately. The optical printing system remains stationary, and a pattern is printed at each die site by moving the wafer with highly accurate displacement motors. This is called the **step-and-repeat** process, and the piece of equipment that provides this action is called a **stepper**.

It is worthwhile mentioning that photoresist is sensitive to ultraviolet (UV) light, which is on the short-wavelength side of the optical spectrum. It does not react to long-

Figure 2.12:
Silicon wafer with
individual die sites

wavelength light, and it can be exposed to yellow light without causing any problems. Because of this, the printing procedure is usually confined to a “yellow-light area” in the fabrication facility. This allows us to see and maintain the operation without ruining the outcome.

Now that we have seen the photolithography process in detail, it is time to introduce some numbers into the discussion. We have talked about defining patterns in general terms, but have not specified actual sizes yet. On-chip feature sizes of modern integrated circuits are specified in units of **micrometers** denoted by μm ; we often call this unit a **micron** for short. By definition,

$$1 \mu\text{m} = 10^{-6} \text{ m} = 1/1,000,000 \text{ m} = 10^{-4} \text{ cm}$$

To get an idea of how small this really is, use a meter stick to see a 1 millimeter (mm) increment; note that $1 \text{ mm} = 10^{-3} \text{ m}$. Next, try to visualize dividing up the 1 mm segment into 1000 pieces; if you can, then each segment will be $1 \mu\text{m}$ long!

The **minimum feature size** of an IC fabrication line specifies the smallest dimension that can actually be transferred to a chip. This is determined by the resolution of the lithographic equipment and the etching processes. Smaller line geometries mean higher circuit density, so that there is a continuous drive to reduce the sizes of chip features. Modern processing facilities can manufacture chips with minimum feature sizes smaller than 1 micron; this defines the **submicron** region of silicon integrated circuits. Advanced state-of-the-art fabrication plants can now produce integrated circuits where the smallest feature is less than $0.1 \mu\text{m}$ wide, driving us into the **deep submicron** region.⁴ The importance of these sizes will be seen when we progress to the details of layout design. Procedures and rules change as we shrink to smaller and smaller processes, and we must ensure that we follow these changes very carefully.

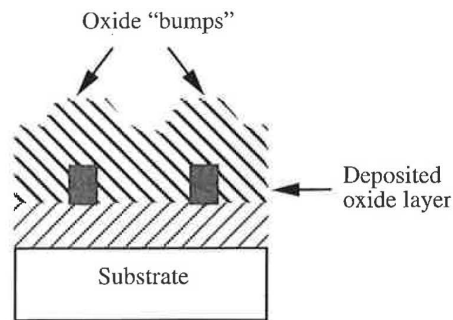
One thing that does not change in the lithography is that a single speck of dust on the photoresist during the exposure step can interfere with the light absorption and destroy the patterning. A single defect can render an entire chip non-functional, so this becomes a severe problem if not dealt with at the beginning. Critical chip-manufacturing steps are carried out in a **clean room** environment where the air is continuously filtered to remove particulate matter. Employees must go through a rigorous cleaning procedure before entering these areas and are required to wear special gowns that cover most of their bodies. In the most critical areas, persons will wear “bunny suits,” whose only direct openings are for the eyes.

2.4 Planarization

Integrated circuits consist of many patterned layers of materials. After a metal layer has been patterned, a layer of insulating oxide will be deposited on top of it, followed by

4. “Deep submicron” usually refers to processes with minimum feature sizes less than about 0.25 microns.

Figure 2.13:
Origin of the
planarization
requirement



another layer of metal. This introduces surface “bumps” in the regions where the metal has been kept, as shown in Figure 2.13. The problem gets increasingly worse with every layer and may make it impossible to reliably form a metal pattern on the upper levels.

Planarization is a general technique developed to provide flat surfaces for every metal layer. The idea is quite simple. After an oxide is deposited over a metal, we microscopically “sand” the surface using various physical and chemical processes to give a flat surface. Once this is accomplished, we can deposit another layer of metal without worrying about the local terrain of the surface. The procedure is summarized in Figure 2.14.

Planarization is mandatory in chips that use more than about two or three metal layers. The most common technique is **Chemical-Mechanical Polishing (CMP)**, which is capable of producing very flat surfaces. All of the process examples that we will examine in this book employ planarization.

2.5 Electrical Characteristics

Now that we have an idea of how patterns are formed, let us examine some of the important electrical characteristics that will affect the design of an integrated circuit.

Figure 2.14:
Planarization

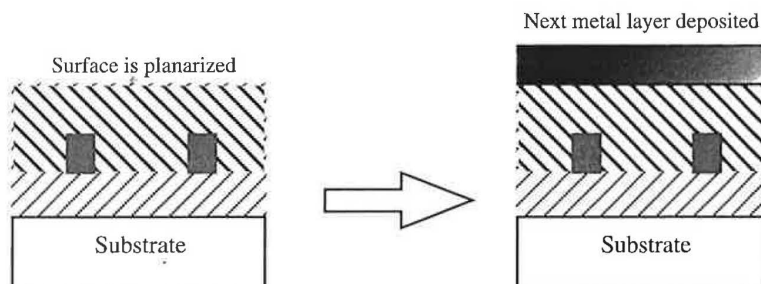
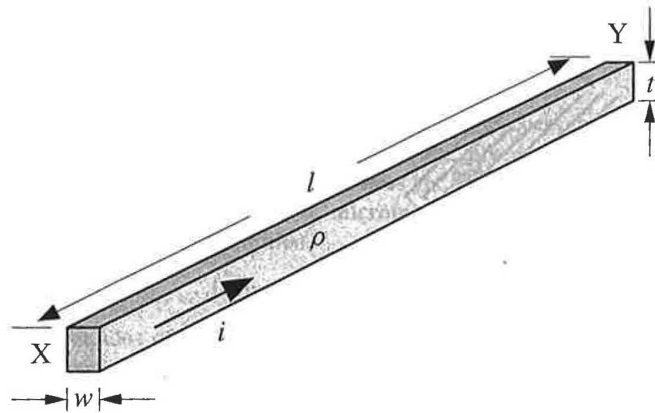


Figure 2.15:
Line geometry



All materials have electrical resistance, R , that resists the flow of electrical current. If we construct a patterned line with the geometry shown in Figure 2.15, then the resistance from X to Y is given by

$$R = \frac{\rho l}{A} \Omega \quad (2.1)$$

where

ρ is the resistivity of the material in $\Omega\text{-cm}$

l is the length of the line in cm

$A = wt$ is the cross-sectional area in cm^2

The resistivity ρ , depends on the material, with conductors such as aluminum (Al) and copper (Cu) having the lowest values. The thickness t of the material is determined by the processing recipe, and the chip designer does not have control over it. The main variables are the width w and the length l . These are also the dimensions seen from a top-view drawing, as portrayed earlier in Figure 2.6(b). The significance of the line resistance is that a current i flowing in the line gives a voltage

$$V_{XY} = iR \quad (2.2)$$

between the front and back. Since voltages are used to represent signals and information, the unwanted (**parasitic**) resistance will affect how well the physical circuit works.

Now note that conducting metal lines are patterned on top of the insulating-oxide layers. This creates a line capacitance C with respect to the silicon substrate. The geometry is detailed in Figure 2.16.

2.6 Silicon Characteristics

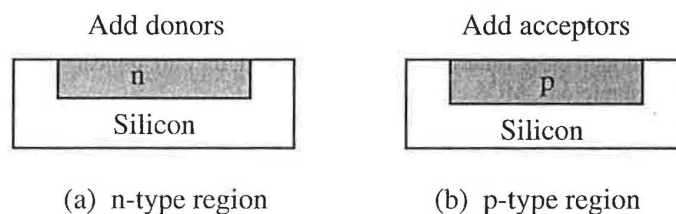
Although we have only looked at metal conducting lines, silicon is the base material for integrated circuits.

A digital integrated circuit is created by using several layers of patterned conducting materials to control the flow of signals on a small piece of silicon, called a **die**. While many layers are metal interconnect that act as microscopic wires, some layers are made with different types of crystal silicon (represented by the elemental symbol Si) which are **semiconductors**. Silicon is a very special type of material that allows us to create a type of electrically controlled switch that can be closed or opened using a control signal. In modern chip design, the switches are called **field-effect transistors (FETs)**. FETs provide the basis for making logical decisions to switch (or not to switch) signals.

Silicon is called a semiconductor because it is a “partial” (or weak) conductor of electricity. It is possible to force a current to flow through silicon, but it only supports relatively small levels, compared to a conductor such as aluminum or copper. Metals can easily handle current flows of amperes, while semiconductor currents are usually measured in units of microamperes ($1 \mu\text{A} = 10^{-6} \text{ A}$) or milliamperes ($1 \text{ mA} = 10^{-3} \text{ A}$). The magnitudes of the current really are not important. What makes silicon (and all other semiconductors) important is the fact that we can add small amounts of other types of atoms to the otherwise pure silicon, which alters how that region conducts electricity. The added impurity atoms are called **dopants**. This results in two distinct “types” of silicon, which in turn, forms the entire basis of microchip design as it allows us to create the FET switching devices.

If we add arsenic (As) or phosphorus (P) in very small amounts to the silicon, we find that the dopants enhance the number of free electrons that are available for conduction. The doped material thus conducts electricity better than it did in a pure silicon state. These dopants are called **donors** because each As or P atom added to the silicon increases the number of free electrons that carry electrical current; in other words, each atom “donates” an electron to the electrical conduction process. Each electron has a negative charge denoted by $-q$ where $q = 1.602 \times 10^{-19}$ Coulombs (C) is the fundamental charge unit. The resulting region of silicon is called **n-type** since it has an excess of negative electron charge carriers. In our drawings, an n-type region is indicated by simply labeling it as shown in Figure 2.19(a).

Figure 2.19:
Labeling of doped
silicon regions



The number of donor atoms added to the sample is represented by the symbol N_d , which has units of numbers of donors per cm^3 ; the strict units of N_d are cm^{-3} . If we let n be the number of free electrons per cubic centimeter (also called the free electron density), then

$$n \approx N_d \quad (2.14)$$

for an n-type material. In this approximation, every arsenic or phosphorus atom donates one free electron to the crystal. If N_d is very large, then the region is labeled as n+; this usually requires N_d to be greater than about 10^{18} cm^{-3} .

The other class of dopants used in silicon processing is elemental boron (B). When boron is added to the silicon crystal, we find that every boron atom adds a mobile positive charge called a **hole** to the conduction process. A hole is the inverse of an electron; it acts like a particle with a charge $+q = +1.602 \times 10^{-19} \text{ C}$, but it still contributes to the conduction process. This creates a **p-type** region and is labeled as shown in Figure 2.19(b). A hole can "absorb" an electron and neutralize it, so we call boron atoms **acceptors** since they can "accept" electrons. It is worth noting that this process is called **electron-hole pair annihilation**, such that

$$1 \text{ electron } (-q) + 1 \text{ hole } (+q) = 0 \text{ Coulombs} \quad (2.15)$$

Denoting the number of boron acceptors per cubic centimeter by N_a , the number of positively charged holes, p , per cubic centimeter is given by

$$p \approx N_a \quad (2.16)$$

This is called a p-type material since there is an excess of positively charged holes in the region. If N_a is very large, then the region is labeled as being p^+ .

A fine point in this discussion is that an n-type material still has a few holes, while a p-type sample always has some electrons. The relationship between holes, n , and electrons, p , is given by the **mass-action law**, which states

$$np = n_i^2 \quad (2.17)$$

where n_i is called the **intrinsic density** and represents the number of electrons or holes in an undoped (pure or intrinsic) sample. In silicon at room temperature ($T = 27^\circ \text{C} = 300 \text{ K}$), $n_i \approx 1.45 \times 10^{10} \text{ cm}^{-3}$. This brings us to an important point:

Electrical conduction in silicon is dominated by either electrons (n) or holes (p) depending upon the material. The resistance of a patterned silicon line is still calculated using equation (2.1). The resistivity of an n-type region is given by

$$\rho_n = \frac{1}{q\mu_n N_d} \quad (2.18)$$

where μ_n is called the **electron mobility** with units of $\text{cm}^2/\text{V}\cdot\text{sec}$. If the material is p-type, then

$$\rho_p = \frac{1}{q\mu_p N_a} \quad (2.19)$$

with μ_p being the **hole mobility**. These equations show that the line resistance varies with the doping densities. In general, the resistivity of silicon regions are much larger than that of metals, so silicon is not typically used for interconnects. The main application of doped-silicon regions is to make transistors.

One important fact from semiconductor physics is that electrons and holes behave differently in the conduction process. Electrons are said to be more mobile than holes. In terms of physical parameters, this means that μ_n is larger than μ_p in any region of a semiconductor. If we create an n-type sample using a donor-doping N_d , and compare it to a p-type sample that has an acceptor doping of $N_a = N_d$, then $\rho_p > \rho_n$. An n-type line thus has a smaller resistance than a p-type line with the same dimensions. This will also be true of the transistors that are created using the two different types of doped silicon.

The capacitance of a line formed by doping a region of silicon is more complicated, since it depends on the bordering material. When n-type and p-type regions touch, a **pn junction** is formed. This leads to a junction capacitance that affects signal delays.

We can now start to see some of the features of the chip structure evolving. Transistors are built in silicon, which forms the base substrate for the entire chip. Alternating metal and insulator layers are deposited on top of the substrate, and these are used for wiring the transistors together. The finished structure represents the physical realization of the circuit in silicon.

2.7 Overview of Layout Design

Chip layout can be characterized as the step in the design hierarchy where an electronic circuit is transferred to a silicon description. The layout designer is responsible for creating the patterns on every layer such that the resulting stacked structure defines

- The electronic switching devices (transistors)
- The wiring that connects the switching devices together

To do this, we must first examine what constitutes a transistor, and then learn the characteristics of the conducting and insulating layers. This results in a set of rules that help guide the complex task of defining every part of every transistor correctly and then providing the metal wiring.

Layout is performed entirely on a computer using a layout editor. Every layer on the chip is patterned so that the resulting layers stack into 3-dimensional structures that constitute the electronic network. A common grid is used throughout the process, and the screen

portrays the entire design, but the patterning information for each layer is stored in a separate database. The data is used to create the reticle so that the patterns can be transferred to the chip during the photolithographic process.

It is the responsibility of the layout designer to insure that the patterns follow the geometrical guidelines established for the fabrication process. Violating the rules may result in a non-functional chip. The layout tools also provide the ability to translate a set of patterns into the equivalent electronic circuit for comparison. The Layout Versus Schematic (LVS) check insures that the patterns accurately represent the desired circuit.

The toolset is used to **extract** a circuit schematic from the layout drawings. This provides a listing of every electronic element and the wiring details; the parasitic resistance and capacitance of every line can also be determined. The extracted file is used to simulate the electronic behavior of the silicon circuit. This simple overview shows that physical design is dependent on every level of the design hierarchy. The shapes and sizes of the material layers created in the layout process determine much of the final electrical characteristics of the fabricated chip. It is therefore considered to be a critical part of the design hierarchy.

2.8 References

- [2.1] Campbell, S. A., *The Science and Engineering of Microelectronic Fabrication*. New York: Oxford University Press, 1996.
- [2.2] Baker, R. J., Li, H. W., and Boyce, D. E., *CMOS Circuit Design, Layout, and Simulation*. Piscataway, NJ: IEEE Press, 1998.
- [2.3] Uyemura, J. P., *A First Course in Digital Systems Design*. Pacific Grove, CA: Brooks-Cole, 1999.
- [2.4] Uyemura, J. P., *Introduction to VLSI Circuits and Systems*. New York: John Wiley & Sons, 2002.

2.9 Exercises

- 2.1 An integrated circuit line has a length of l_o and a time-constant delay of t_o . Find the delay in terms of t_o for lines of the following lengths.
 - (a) $l = 2l_o$
 - (b) $l = 4l_o$
 - (c) $l = 8l_o$

- 2.2 A metal line is $0.60\text{ }\mu\text{m}$ thick and $0.25\text{ }\mu\text{m}$ wide. It has a length of $25\text{ }\mu\text{m}$ and runs over an oxide layer that is $1.2\text{ }\mu\text{m}$ thick.
- (a) The resistivity of the material is known to be $\rho = 1.8\text{ }\mu\Omega\text{-cm}$. Calculate the line resistance.
 - (b) Find the line capacitance in farads (F) if the permittivity of the oxide is $\epsilon_{\text{ins}} = 3.54 \times 10^{-13}\text{ F/cm}$.
- 2.3 A silicon region is a doped n-type with $N_d = 2 \times 10^{18}\text{ cm}^{-3}$.
- (a) Calculate the electron density n .
 - (b) Find the resistivity of the region if $\mu_n = 125\text{ cm}^2/\text{V-sec}$.
 - (c) A patterned line is made in the n-type region. The width is $0.30\text{ }\mu\text{m}$ and the line is $9.6\text{ }\mu\text{m}$ long. The layer is $0.15\text{ }\mu\text{m}$ thick. Find the resistance.
- 2.4 A silicon region is a doped p-type with $N_a = 2 \times 10^{17}\text{ cm}^{-3}$.
- (a) Calculate the electron density p .
 - (b) Find the resistivity of the region if $\mu_p = 80\text{ cm}^2/\text{V-sec}$.
 - (c) Find the line resistance if a line is made in the p-type region with the same dimensions as in Problem 2.3(c).

CMOS Technology—A Basis for Design

CMOS technology is the driving force behind high-density chip design.¹ It is safe to say that most of the integrated circuits in your computer are CMOS-based. The *MOS* part of the acronym stands for **metal-oxide-semiconductor** and originates from the basic layering scheme in the original structures developed in the 1960s. The *C* stands for **complementary** and originates from the circuit-design methodology. In this chapter, we will examine the processing flow used to fabricate CMOS chips and learn how to create transistors. This will provide the basis for our detailed introduction to physical design in Chapter 4.

3.1 Meet the MOSFETs

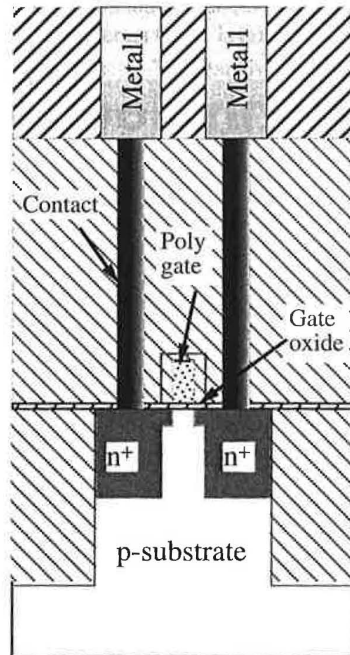
MOSFET is an acronym for **MOS field-effect transistor**. It is the smallest switching element that can be used on a mass-produced integrated circuit, so that much time, money, and energy has been spent designing and building smaller and faster MOSFETs. In our treatment, a MOSFET is simply a 3-dimensional structure built by sandwiching patterned regions from different material layers. Unlike the metal wiring we studied in previous sections, transistors are fabricated directly in the silicon.

There are two **polarities** or **types** of MOSFETs found in CMOS. An n-channel MOSFET conducts electrical current using negatively charged electrons. These will be called nMOS transistors or simply nFETs in this discussion. A p-channel MOSFET (pMOS or pFET) is exactly the opposite: it conducts electrical current using positively charged particles called holes. nFETs and pFETs are said to be electrical complements of each other, and they are often used in pairs (hence the origin of the complementary in CMOS).

Let us examine an nFET first. A cross-sectional view is shown in Figure 3.1 and is similar to the figures you will see using the Microwind 2-dimensional view command.

1. CMOS is pronounced as *see moss*.

Figure 3.1:
nFET cross-section



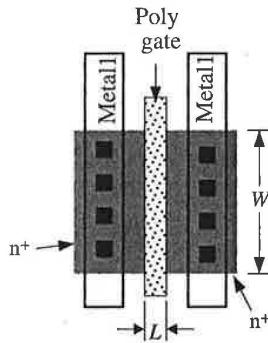
Cross-hatched areas in the drawing represent insulating oxide regions. The transistor itself is built in a silicon that is denoted by “p-substrate” in the drawing. The “p” designation means that small amounts of boron (B) have been added to the silicon crystal. This increases the number of positive charges in the region, giving rise to the designation “p-type” material.

The central region of the nFET is the most important. The small region marked “poly gate” is the control electrode for the device. As we mentioned earlier, “poly” is a shorthand designation for poly-crystalline silicon, which is just a conducting layer to us. The voltage on the gate determines the ability of the nFET to conduct an electrical current between the two separated “n⁺” regions in the silicon. An n⁺ region is one where we have a large excess of negatively charged electrons to carry an electrical current. These are created by adding arsenic (As) or phosphorus (P) to the otherwise pure silicon.

The contacts in the drawing represent metal filler that provides an electrical connection between the n⁺ regions in the silicon and the Metal1 patterns at the top of the structure. They are used to gain access to the silicon-level device from the metal layers above the chip and are created by etching holes in the oxide layer, then adding the metal. Tungsten (W) is probably the most common contact material used with this approach. Copper (Cu) can also be used for contacts, but the procedure is quite different.

The last, but perhaps most important, feature of the nFET is the gate oxide between the poly gate and the silicon surface. This forms an insulating layer between the conducting poly gate and the p-substrate. In practice, the p-substrate is connected to an electrical ground, so that applying a voltage to the gate causes the structure to act like a capacitor.

Figure 3.2:
nFET top view



The gate oxide is very thin. Denoting the thickness by t_{ox} , a typical value in a submicron process is $t_{ox} = 30 \text{ \AA} = 3.0 \text{ nm}$, where $1 \text{ \AA}$ (Angstrom) $= 10^{-10} \text{ m} = 10^{-8} \text{ cm}$ and 1 nm (nanometer) is 10^{-9} m . Compared with a poly thickness of about $0.2\text{--}0.5 \text{ }\mu\text{m} = 200\text{--}500 \text{ nm}$, it is really too thin to properly scale in the drawing. In early-period MOSFETs, the gate was made out of aluminum metal, not poly as in modern transistors. Combining this with the gate oxide over the semiconductor substrate gives the origin of the MOS acronym: metal on oxide on semiconductor. Even though metal gates are no longer used in practice, MOS is still used to describe this capacitor system. We will study it in more detail when we introduce the electrical characteristics of MOSFETs.

Layout of a physical design is based on the structures as seen from a top view. An nFET is drawn in Figure 3.2. No oxide layers are shown explicitly, and the entire device is assumed to be surrounded by an oxide barrier as in the cross-sectional view. Contacts are shown as black squares; note that there are several on each side of the transistor. Two new parameters have been defined. The **channel width** is denoted by W and shows the extent of the n^+ region parallel to the gate. It has practical units of microns (μm). The **channel length** L is equivalent to the width of the poly gate line, and also has practical units of microns. Both W and L are called “drawn” values because they represent the mask values that we draw on the screen. For most transistors, L will be the same as the minimum feature size. After the device is fabricated, the final (effective) values will be slightly different. The aspect ratio of the device is defined by W/L and is an important electrical design parameter that is set in the layout.

The nFET circuit symbol is shown in Figure 3.3. The symbol is used to represent the n-channel MOSFET in electrical circuit schematic diagrams. It shows the basic architectural features of the device, making it easy to translate between the symbol in circuit schematics and the physical structure. The poly gate is the control electrode that is electrically

Figure 3.3:
nFET circuit symbol

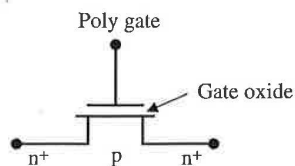
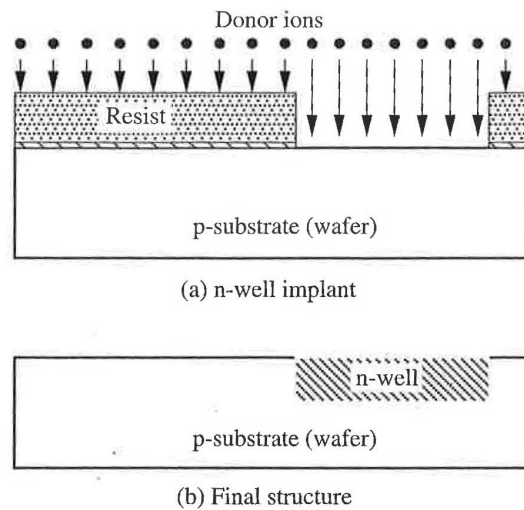


Figure 3.7:
Formation of the
n-well region for
pFETs



The formation of the well “layer” is shown in Figure 3.7. A mask is used to define the n-well regions. After exposure and development, a hardened layer of photoresist covers and protects all sections of the silicon surface that are not n-well. An n-type ion implant (arsenic or phosphorus) can penetrate unprotected areas and form the n-regions as portrayed in Figure 3.7(a). The resulting structure in Figure 3.7(b) shows the n-well after the annealing step is completed.

Mask 2: Active Areas and FOX

High-density integrated circuits need to pack transistors very close together on the silicon surface. To insure that adjacent devices do not have any unwanted electrical conduction paths among them, individual transistors are surrounded by glass (oxide) regions that are created in the silicon. This is a general form of device **isolation**.

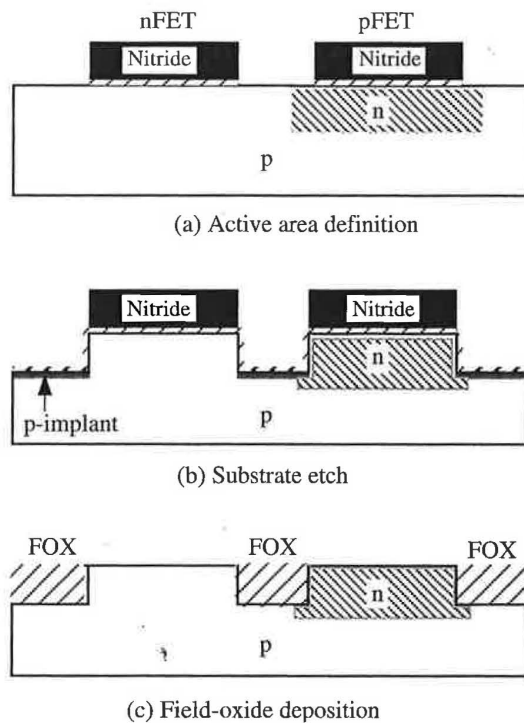
Regions of a chip are divided into two categories, depending upon their usage. Transistor sections are called **active** areas, since they contain active electronic devices. Any space that is not active is defined to be a **field** region. Symbolically, we may write that

$$\text{Active} + \text{Field} = \text{Chip surface}$$

Field oxides (FOX) are formed to provide device isolation.

Isolation regions are defined by the **active** area mask. This mask is used to define FET regions on the surface by patterning a dual layer of silicon nitride (Si_3N_4) and oxide, as shown in Figure 3.8(a). Nitride is used because it is a very dense dielectric that protects the silicon surface. The nitride/oxide layer itself is used as a mask for the etching process depicted in Figure 3.8(b). RIE is used to form trench-like regions in the silicon to physically separate neighboring transistors. A thin layer of oxide is then grown over the surface, followed by a p^+ implant into the trench. Depositing oxide on top then fills the field

Figure 3.8:
Active area
definition and
FOX formation

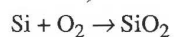


regions, yielding the view in Figure 3.8(c). The key to this type of device isolation is to insert a glass barrier between neighboring devices.

At this point, separate ion implants are made to nFET regions and pFET regions to adjust the “turn-on” parameter called the **threshold voltage**. We will examine the effect of this in later chapters.

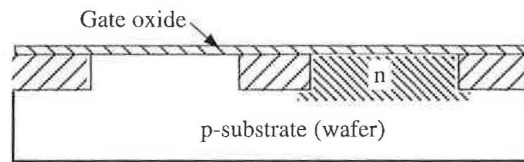
Mask 3: Polysilicon Gate

The next sequence of steps is directed towards forming the actual transistors. This starts with the growth of the gate oxide on the surface. Passing oxygen over a heated silicon wafer forms silicon dioxide according to the reaction

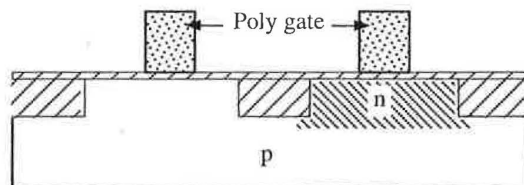


The oxide is overgrown to insure that there are no “pinholes” in it, and then etched back to a working thickness of t_{ox} , as in Figure 3.9(a). No mask is required for this step, since the oxide is grown over the entire surface of the wafer. Some processes allow the user to selectively block the etchback, which results in FETs with thicker gate oxides. These yield devices with higher threshold voltages that can be used with higher power-supply levels.

Figure 3.9:
Gate oxide growth
and poly gate
formation



(a) Gate oxide growth



(b) Poly gate formation

MOSFETs are formed using what is called a **self-aligned** technique where the gate and n^+ (or p^+) regions automatically are aligned to each other. The gate is formed by depositing doped silicon on top of the gate oxide and then patterning the layer to define the gates; Figure 3.9(b) shows the resulting cross-section. Silicon has a natural tendency to form a crystal, but silicon dioxide (the gate-oxide layer) is amorphous, so it does not provide a good base. Because of this, the silicon forms small regions of crystals, and the overall material is called **polycrystalline** silicon, or **poly** for short. For modeling purposes, we usually treat it as being similar to regular silicon regions.

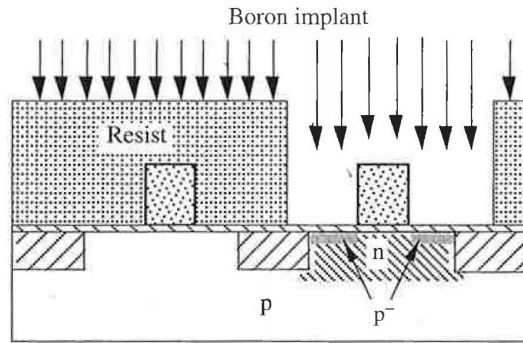
Masks 4 and 5: n^+ and p^+ Doping

The next steps are ion implants that are used to define the p^+ and n^+ regions of pFETs and nFETs, respectively. The “+” designation of p- and n-type regions is used to imply that the doping density is very large. In device terminology, the regions form what are known as the **drain** and **source** terminals of a FET, so that they are also called the drain/source masks. In this sequence, we will create what is known as a **lightly doped drain (LDD)** MOSFET. These structures are used to reduce some small device effects due to very energetic particles called **hot electrons** and **hot holes**.²

Let us start with the p-type boron implant use to form pFETs. The first step is a “light” boron implant that creates p^- (lightly doped) regions in the pFET, as shown in Figure 3.10. The resist mask blocks the boron from entering nFET sites and concentrates the dopants over pFET regions. The process is called “self-aligned” because the poly gate itself acts as a mask to the incident ions, blocking them from entering the silicon below. This gives p^- silicon doping on the left and right side of the gate, but the silicon directly under the gate is still n-type. An nFET structure is created in the same manner. A mask is used to create a

2. In this context, “hot” refers to a particle’s energy as related to a thermodynamic temperature.

Figure 3.10:
First p-type
pFET implant



resist block over pFET regions while providing access to nFET sites. Ion implanting donors (arsenic or phosphorus) creates lightly doped n^- drain and source regions, as shown in Figure 3.11.

LDD MOSFETs combine the lightly doped drain and source regions described above with heavier p^+ and n^+ sections that make the material more conductive. Fortunately, the next steps do not require a separate mask, making the process very attractive from an economic viewpoint. The key to finishing the LDD structure is to first create **spacer** oxides on the poly gates. The sequence is shown in Figure 3.12. Oxide is deposited over the surface of the wafer as illustrated in Figure 3.12(a). The oxide is then etched vertically downward. From the top, the oxide on the sidewalls of the poly gates is thicker than other regions, so it doesn't etch all the way down. The remaining sidewall of oxide shown in Figure 3.12(b) are the spacers.

The heavy drain/source doping is completed using the spacers and poly gates as masks. Figure 3.13 portrays both the pFET [in (a)] and nFET [in (b)] implants and resulting structures. This step completes the formation of the transistors.

Figure 3.11:
Initial n-type
nFET implant

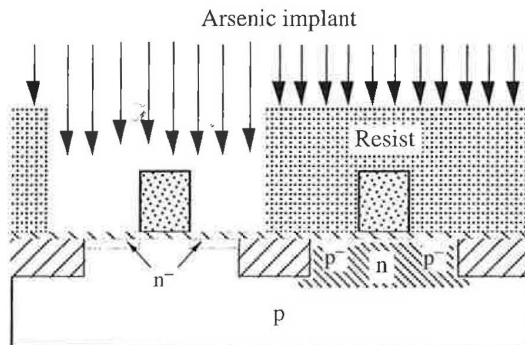
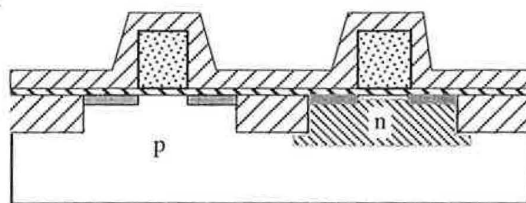
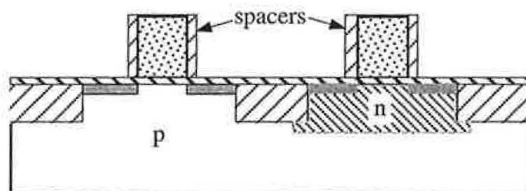


Figure 3.12:
Spacers

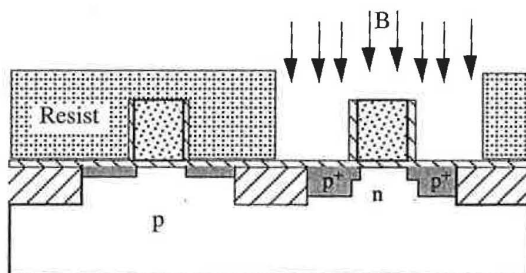


(a) Oxide deposition

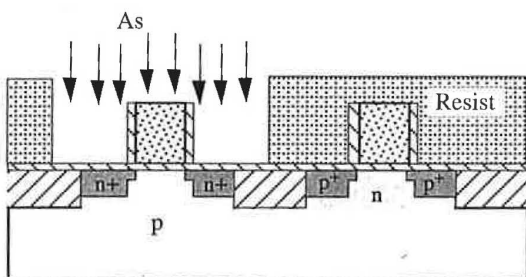


(b) Etchback and spacer formation

Figure 3.13:
Deep p⁺ and
n⁺ FET implants

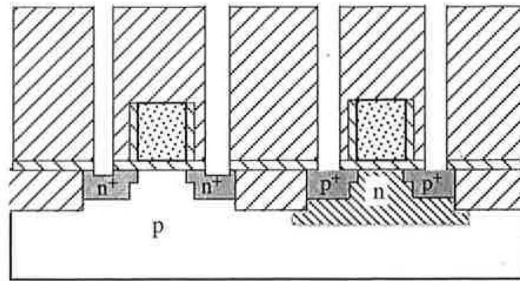


(a) Deep p⁺ implant

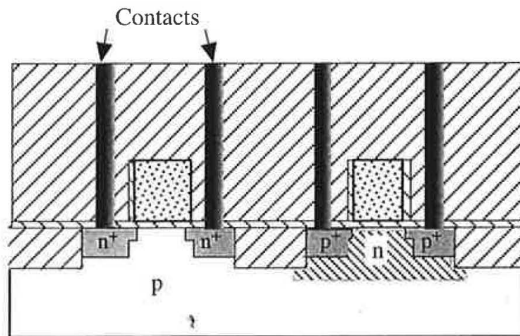


(b) Deep n⁺ implant

Figure 3.14:
Oxide cuts
and plugs



(a) Oxide deposition and etching



(b) Contact plug fill

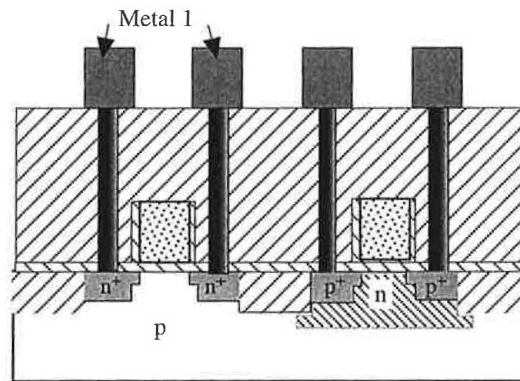
Mask 6: Oxide Contacts

Once the FETs are fabricated, we must provide a means to wire them into the circuit. This is accomplished by alternate oxide-metal layerings, where the metal layers are patterned to provide the signal flow paths. The first step is shown in Figure 3.14(a) where an oxide layer is deposited uniformly on the FETs, and contact cuts (holes) are patterned in the insulating material. These are filled with conducting plugs, yielding the cross-section in Figure 3.14(b).

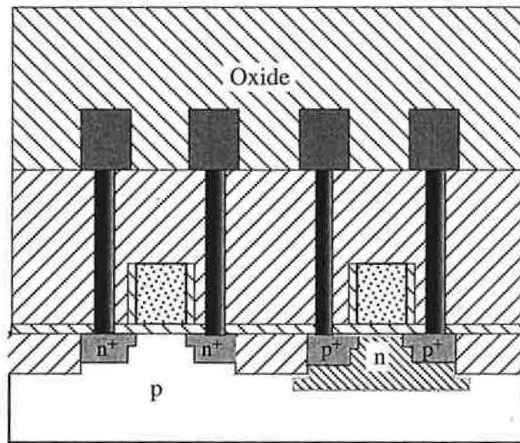
Mask 7: Metal1

The Metal1 layer is deposited on top of the oxide and then patterned, giving us the structure illustrated by Figure 3.15(a). This serves as the base for higher metal layers that are added by continuing the process. The next step is to deposit an oxide layer on Metal1 and then planarize the top as shown in Figure 3.15(b).

Figure 3.15:
Metal1 oxidating



(a) Metal1 deposition and patterning

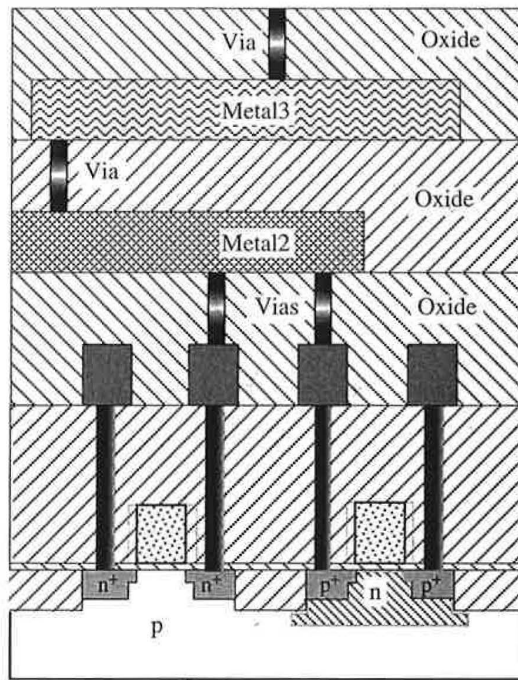


(b) Oxide and planarization

Masks 8+: Vias and Higher-Level Metal Layers

The oxide on top of Metal1 is then patterned with contact holes and filled with plug material to allow electrical contact to higher metal layers. Connections between metals are usually called **vias**. The next metal (Metal2) is then deposited, patterned, and covered with oxide. The oxide over Metal2 is patterned and then is itself covered with oxide. Continuing this process gives a cross-section like that shown in Figure 3.16.

Figure 3.16:
Higher-level metal
layers and vias

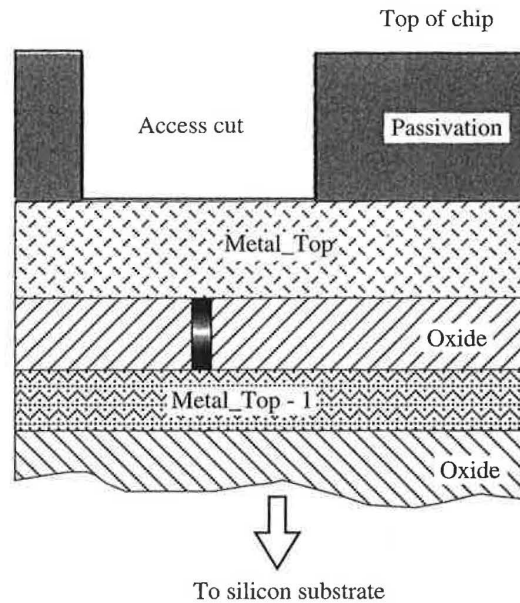


Final Masks: Metal_Top and Passivation

Let us call the top conducting layer Metal_Top. It is sometimes made extra thick, particularly if integrated on-chip inductors are included in the process capabilities. Metal_Top is covered with a layer of silicon nitride that is used as the **passivation** coating. Nitride is a very dense dielectric that protects the chip from contaminants in the atmosphere. This is important for reliability considerations, as foreign substances may lead to chip failure if they diffuse into the structure. The passivation mask is used to open holes in the nitride that give electrical access to the top metal layer. In other words, passivation cuts define the entrance and exit points to the outside world. Figure 3.17 shows the Metal_Top and passivation coatings. This completes the chip fabrication sequence.

After the wafer is through the process cycle, it is put into a testing apparatus that probes each die. Testing is a very important, but expensive, part of the manufacturing sequence. Every die is powered up and subjected to various input combinations whose results can be observed at specified outputs. Not every die is functional or meets the specifications. Those that fail are marked and eventually disposed of. When the wafer test is

Figure 3.17:
Top metal and
passivation layer



finished, a diamond-tipped blade is used to score the wafer between the die sites. A small amount of pressure cracks the crystal along the score lines, and individual die can then be mounted in packages. Additional testing may be required for the packaged devices.

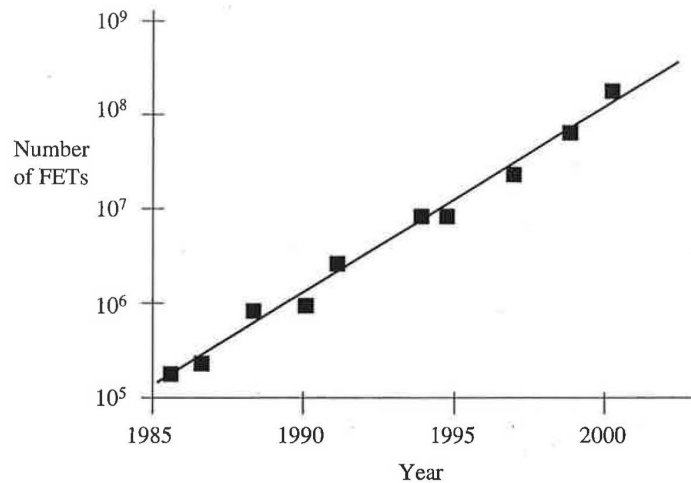
3.3 Submicron CMOS Processes

CMOS processing has advanced at a remarkable rate in the past few years. Minimum feature sizes smaller than $0.1\ \mu\text{m}$ are common, and new advances appear poised to continue the trend of the “incredible shrinking MOSFET.” Many factors contribute to this trend.

3.3.1 Moore’s Law and Integration Levels

Early integrated circuits contained only a few transistors and other components. These were used to create basic logic functions, such as the NOT, NAND, and NOR; and early ICs were placed in metal or plastic packages with only a few input/output (I/O) pins. Modern chips have several hundred million transistors, and the number of FETs will exceed one billion in the near future. Advanced IC packages have several hundred pins, yet engineers still complain that they are “pin-limited” in their designs. Of course, this massive change didn’t take place overnight.

Figure 3.18:
Example of
Moore's Law



Gordon Moore, one of the founders of Intel Corporation, made a prediction that has since become known as "Moore's Law." It states that

The number of transistors on a die will double every 18 months.

At the time of his statement, the integrated circuits industry was experiencing moderate growth, but even then it seemed very optimistic. Few believed that it would extend into the 21st century. But it has! Technologists often marvel at the fact that Moore's Law, which implies shrinking dimensions with a constant rate of process improvement, still seems to be in effect today. Figure 3.18 shows an example of Moore's Law for some representative chips selected from the literature. The continuous increase in the device count is easily seen.

The concept of **integration levels** was introduced to describe the increase in transistors and functionality over the years. Traditionally, chips have been divided into four categories, although the borders are not well defined. These are given by the acronyms below; note, however, that the number of transistors quoted in each level is merely a rough estimate.

- **SSI** (Small-Scale Integration) with less than about 1,000 transistors. SSI includes chips with individual logic gates.
- **MSI** (Medium-Scale Integration) to around 100,000 transistors. An example of an MSI chip is a 4-bit arithmetic logic unit (ALU) or a basic calculator.
- **LSI** (Large-Scale Integration) to around 1,000,000 transistors. 8- and 16-bit microprocessors and basic digital signal processors (DSPs) are in this category.
- **VLSI** (Very Large-Scale Integration) to around 100,000,000 transistors. This includes the current generation of microprocessors that have about 40–50 million transistors.

Current day chip design is at the VLSI level. For circuits with a billion or more FETs or logic gates, the next level is

- **ULSI** (Ultra Large-Scale Integration) with about one billion transistors, which some have coined
- **GSI** (Giga-Scale Integration) as an alternate.

CMOS VLSI has been the driving force in computer hardware, digital signal processing, telecommunications, and a host of other fields.

3.3.2 Yield

As mentioned in the previous section, not every die on a wafer is functional. Wafers are usually processed in large groups called lots. Every wafer in the lot is subjected to the same processing steps, so it is meaningful to trace a lot through the fabrication sequence to see how good the processing is.

The **yield** Y of a process is defined as

$$Y = \frac{\text{Number of good die}}{\text{Total number}} \times 100 \quad (3.1)$$

and gives the total percentage of good die. Obviously, a high yield is desirable as each bad die subtracts from the revenue. When a new chip is put into production, the yield may be very low because the design needs to be improved. However, as the design is modified and the recipes are refined, the yield increases to a point where it is profitable to manufacture and sell it.

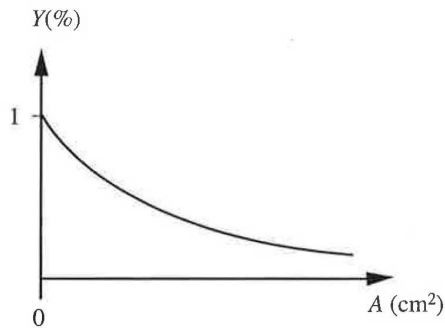
Many factors contribute to the yield. One of particular interest to us is the fact that the yield decreases with increasing die area. The simplest estimate to this dependence is

$$Y = e^{-DA} \times 100\% \quad (3.2)$$

where A is the die area in cm^2 and D is called the **defect density** in units of defects per cm^2 . The defect density describes point defects on the wafer surface and is significant because a single defect can ruin the circuitry on a die. Typical values of D are less than about one defect per cm^2 . Figure 3.19 shows a plot of yield Y as a function of die area A . This shows that smaller die areas result in higher yields. While better yield analysis is possible, this rule generally remains valid.

Increased yield is one factor that drives the industry to smaller and smaller feature sizes. Decreasing the size of a chip results in a higher profit margin and also allows more complex systems to be integrated.

Figure 3.19:
Yield as a function
of die area



3.4 Process Technologies in Microwind



Microwind has been structured in a manner that allows different processes to be accessed and studied. This is accomplished using the selection of a **foundry** for a design.

In semiconductor terms, a foundry is a central chip-fabrication plant that accepts designs from the outside world and uses the design files to manufacture the chips. This allows design companies to create proprietary chips without having to invest several billion dollars into a manufacturing facility. In educational circles, access to various foundries is most easily accomplished using MOSIS.³ This organization was originally established by the U.S. government for the purpose of providing government contractors and universities with a means to fabricate test chips.

Foundries characterize different process lines by the minimum allowed feature size. Other specifications are also involved, such as the addition of another polysilicon layer (Poly2) or the ability to do high-speed analog circuits. However, the minimum resolution is usually the key parameter for a process. It generally implies the smallest dimension for the drawn channel length L in a MOSFET and has a great effect on the characteristics of the circuit. All pattern dimensions are scaled with the feature size, so that the interconnect wiring provided by metal layers depends upon the specification.

To see the different foundry choices available in Microwind, launch the program and execute the Menu command sequence

File ⇒

Select Foundry

as shown in Figure 3.20. The keyboard sequence **[Ctrl][F]** may also be used. Either brings up the Foundry window shown in Figure 3.21. This lists several different options that may

3. MOSIS is an acronym for MOS Implementation Service.

Figure 3.20:
Foundry access

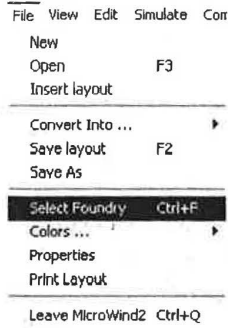
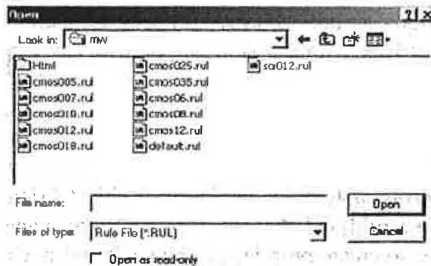


Figure 3.21:
Foundry selection
menu



be used as a design basis.⁴ The file names are written so as to imply the minimum feature size of the process. For example,

cmos012.rul

specifies a 0.12 μm minimum feature size, and it is the default technology that is loaded when Microwind is launched. The file

cmos025.rul

is for a 0.25 μm process. The default can be set to any technology.

It is important to remember that the number of metal layers varies among the different processes. The 1.2 μm process has only Metal1 and Metal2 interconnect, while the 0.25 μm design can access 6 metal layers. If you select a specific technology file, you will be returned to the original screen and the technology and number of metal layers will be listed below the work area. Shrinking the feature size corresponds historically to improved technologies, and there are many other differences that will be found as one progresses from the 1.2 μm to the deep submicron levels.

4. The actual list of processes may be different, depending upon the version of Microwind that you have.

In a realistic design environment, scaling a process down to smaller geometries is much more involved than just reducing the feature size. Electrical parameters change, as do the characteristics of the FETs. And some effects that weren't noticeable in "big" geometries become major problem areas as the dimensions are reduced.

A chip designer must pay attention to these differences as the technology advances. Our philosophy will be to provide a general background in chip layout and simulation that can be extended to various technologies. We will always specify the technology that we are working in.

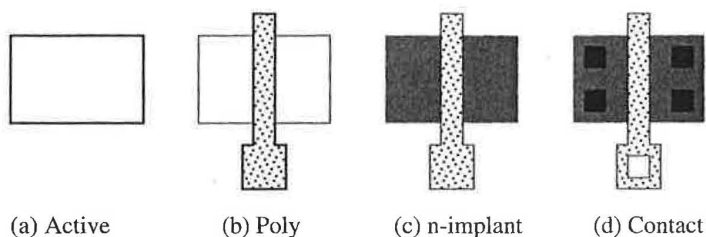
3.5 Masks and Layout

The process flow described in the previous section illustrates the relationship between the mask patterns and the layer sequence. Physical design revolves around designing a set of masks that result in 3-dimensional substructures, such as FETs, and the associated wiring on the chip.

The masking order in the basic process was as follows:

- n-well
- Active
- Poly
- p-implant
- n-implant
- Contact
- Metal1
- Via
- Metal2
- Via
- :
- Metal_Top
- Passivation

where oxide layers are deposited and patterned by the Contact and Via masks. Chip layout requires that every mask be patterned, so that the 3-dimensional layered circuit creates all of the transistors and wires them together correctly. Although this may sound somewhat complicated, it is simplified by the observation that masks can be divided into two categories: FETs and interconnect. This allows us to study the two separately.

Figure 3.22:
nFET masking
sequence

3.5.1 FET Structures

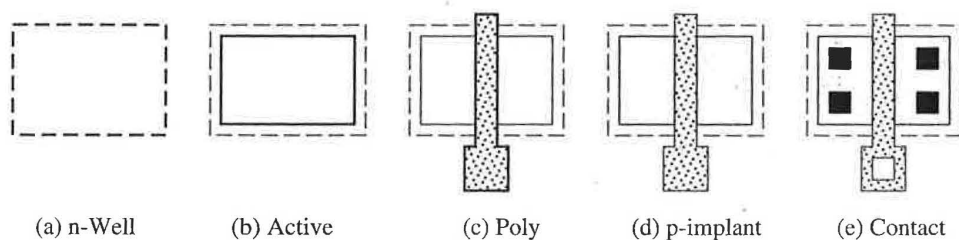
Three-dimensional MOSFET structures are formed using distinct masking groups. Reviewing the processing sequence shows that the following patterned layers are needed for an nFET:

- Active
- Poly
- n-implant

This sequence is illustrated by the top-view drawings in Figure 3.22. This shows how individual masks are aligned to create a 3-dimensional stack with the nFET layering structure. The sequence for a pFET is similar, and consists of the masking steps:

- n-well
- Active
- Poly
- p-implant

The top-view masking sequence for a pFET is shown in Figure 3.23. Note that we have included contacts to the n^+ / p^+ regions and the poly gate to allow for connections to the rest of the circuit. In both devices, the drawn channel width W is set by the vertical dimension

**Figure 3.23:** pFET masking sequence

of the Active mask, while the drawn channel length L is the given by the horizontal extent of the Poly mask.

We can generalize the approach symbolically by defining a masking layer in the abstract form

$$\text{Mask_Pattern} = \text{Layer} + \text{Not_Layer}$$

In this expression, **Layer** represents the geometrical features that are created by the mask. For example, **Active** would denote the collection of all active areas on the chip, while **Not_Layer** are the field regions. Similarly, **Poly** represents the presence of a poly line. Using set notation,

$$(\text{Layer1}) \cup (\text{Layer2})$$

indicates the **union** of features where either **Layer1** or **Layer2** patterns exist. This is logically equivalent to the OR operation. The **intersection**

$$(\text{Layer1}) \cap (\text{Layer2})$$

is where patterns from both layers exist, and is equivalent to the AND logic operation.

Using this symbology, we can define regions by masking layers in a compact manner. We start with doped semiconductor sections. An n^+ region will be denoted as **ndiff** and is created by

$$\text{ndiff} = (\text{Active}) \cap (\text{N-implant}) \cap (\text{Not_Poly})$$

A p^+ region is labeled as **pdiff**, and is created by

$$\text{pdiff} = (\text{Active}) \cap (\text{P-implant}) \cap (\text{Not_Poly})$$

Note that **ndiff** and **pdiff** are derived from other masking layers.⁵ Derived masks do not need to be designed separately since they can be created from existing data. By these definitions, both **ndiff** and **pdiff** are created in active regions of the p-substrate. To build an nFET, we combine the masks as

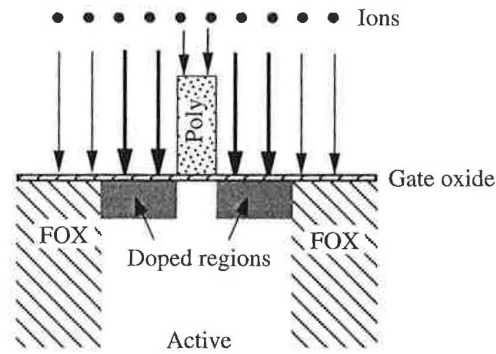
$$\text{nFET} = (\text{ndiff}) \cap (\text{Poly}) \cap (\text{Contact})$$

where we have included the contact opening for completeness. Similarly, a pFET is

$$\text{pFET} = (\text{N_well}) \cap (\text{pdiff}) \cap (\text{Poly}) \cap (\text{Contact})$$

5. The terms "ndiff" and "pdiff" are short for n-diffusion and p-diffusion, respectively. This terminology originates from early silicon fabrication schemes where the n^+ and p^+ regions were doped by diffusing impurity atoms into the silicon substrate using heat. Ion implantation has replaced impurity diffusion, but the terminology still remains (like MOS).

Figure 3.24:
Creation of a self-aligned transistor



This places all pFETs in n-well regions as required in the process flow. These types of expressions help us visualize how features of each layer stack to form the 3-dimensional devices on the finished chip.

A very useful layout rule to remember arises from the self-aligned gate process. The implant step in Figure 3.24 details the formation of a doped region. Ions are incident, or impinge over the entire surface of the wafer, but can be blocked from entering silicon by either the poly gate layer or field oxide (FOX). The only sections that are doped are those where a layer of thin gate oxide is the only barrier that the ions must penetrate. This observation is used as the basis for the general statement that

Poly drawn over diffusion creates a FET.

In terms of the terminology introduced above, we can be a bit more specific and say

Poly over ndiff gives an nFET.

Also,

Poly over pdiff in n-well gives a pFET.

These rules help us remember how to design FETs in layout drawings. It also establishes some rules of layering in that these are always true, so that you must be careful when drawing poly patterns around ndiff and pdiff if you do not want to inadvertently form a transistor.

3.5.2 Interconnects

The remaining masking layers in the basic CMOS process are dedicated to wiring the circuits together. The conducting layers are metals with insulating oxides inbetween. The important masks are:

- Contact
- Metal1
- Via
- Metal2
- Via
- :
- Metal_Top

Metal masks provide the wire routing, while contacts and vias allow us to move between conducting layers to make the routing problem easier.

Routing requirements are based on the design of the circuits that are implemented on the chip. Some are simple, while others can be quite complicated. As we will see later, interconnect wiring differs greatly as we move from small, simple circuits to large, complex systems.

3.6 The Microwind MOS Generator



Now that we are familiar with the CMOS fabrication sequence, let us use Microwind to create our first transistor layouts.

Launch Microwind and examine the screen. The Palette window appears. The top portion of the Palette menu is shown in Figure 3.25. Notice the button in the lower-left side of the screen shot that looks like a FET circuit symbol. This provides access to the MOS generator routine in Microwind. Simply place the cursor on the button as shown and left-click the mouse to bring up the dialog screen. Alternately, you can use the menu commands

Edit ⇒
Generate⇒
nMOS

which brings up the same screen.

Figure 3.25:
MOS Layout
Generator button

