CHAPTER

1

IC Fabrication Overview

SECTION 1

Introduction

Chapter 1 contains a brief synopsis of the detailed discussion to come. It serves as an orientation and reference to give the reader a sense of the big picture and how much goes into the making of a "chip."

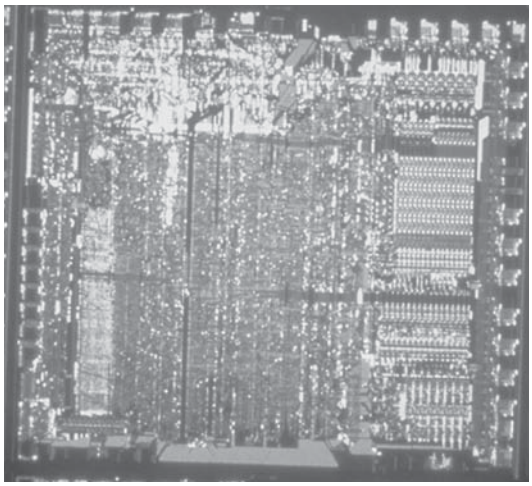


Figure 1-1: Example of an Integrated Circuit (Microprocessor)

1.1 Integrated Circuits

An integrated circuit (IC) is a miniaturized electronic device made by combining discrete electronic components into a single device, all manufactured as a whole. The IC may contain millions of transistors, diodes, resistors, capacitors, and other components integrated into a single chip.

Today, the term "chip" is often the favorite name given to ICs and will be used extensively throughout this discussion.

The IC was originally called the *monolithic* integrated circuit. "Monolithic" means "single-stone," referring to the fact that the entire circuit is built on—and in—a single piece of "stone," the semiconducting material. In 1958, Robert Noyce (one of the founders of Intel), then at Fairchild Semiconductor and Jack Kilby of Texas Instruments were jointly awarded the patent for the IC, even though they did the work separately and at different companies. Noyce built his device on silicon and Kilby built his on germanium.

The important difference in the two approaches to making the first IC was in the way they were wired together. Kilby used thermo-compression wire bonding to connect the components; Noyce used a patterning process and formed the wires by etching a thin aluminum film on the surface of the wafer. Noyce's technique became the standard method.

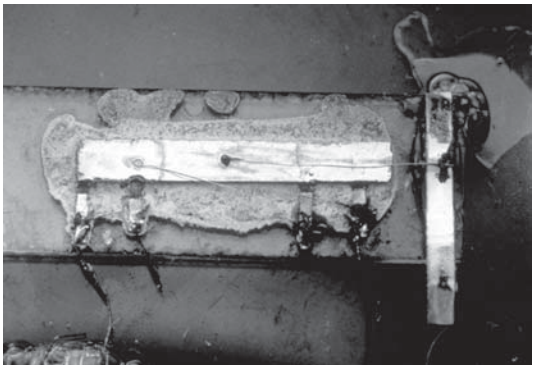


Figure 1-2a: Kilby's First IC

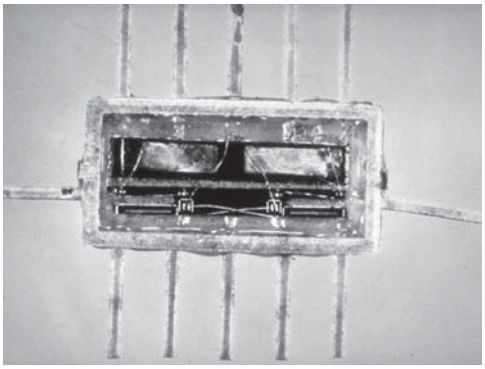


Figure 1-2b: Commercial Version of TI IC

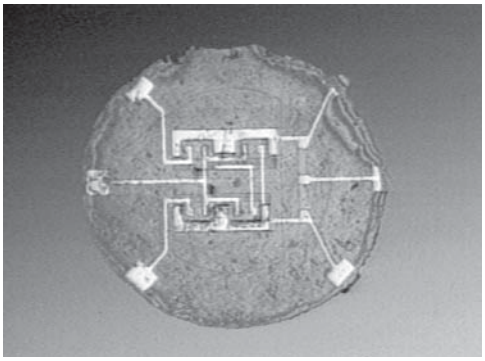


Figure 1-2c: First IC Using Thin Film Wiring (Noyce/Fairchild Semiconductor)

Silicon (Si), the second most abundant element on earth, is the most familiar semiconductor material as well as the most widely used. Silicon technology is the subject of this book. The vast majority of all the ICs produced in the world today are made using silicon.

Germanium (Ge) is another familiar semiconductor. The first transistor was made of germanium and it was the principle material used for the first decade of the industry. It is still is of interest today.

Silicon and germanium are elements, but many semiconducting materials are chemical compounds formed between two or more chemical elements. Not surprisingly, they are called *compound semiconductors*. Gallium arsenide (GaAs), gallium nitride (GaN) and indium phosphide (InP) are examples of compound semiconductors. Although they are not as well recognized, they play a large role in several applications such as high frequency devices used in cell phones and in optoelectronics, most notably in fiber optics applications (telecommunications applications using glass fibers instead of copper wires). Compound semiconductors are also found in applications such as light emitting diodes (LEDs), which are used for traffic lights; they save energy and reduce maintenance costs, as well as producing more intense light.

Semiconductor materials make the magic happen for ICs. Pure semiconductors are not very good conductors of electricity. However, they have a quite handy property that allows them to become relatively good conductors if special impurities, called *dopants*, are added. Doped semiconductors possess either an excess of mobile negative charges (n-type), or an excess of (apparently) mobile positive charges (p-type) depending on the type of dopant chemical that is added. The mobile charges in n-type material are electrons. The mobile charge in a p-type conductor is called a *hole* and it is treated as though it were an actual positive charge carrier. The physics of semiconductor devices is very interesting, and the reader will find some good suggestions for further reading in the Bibliography.

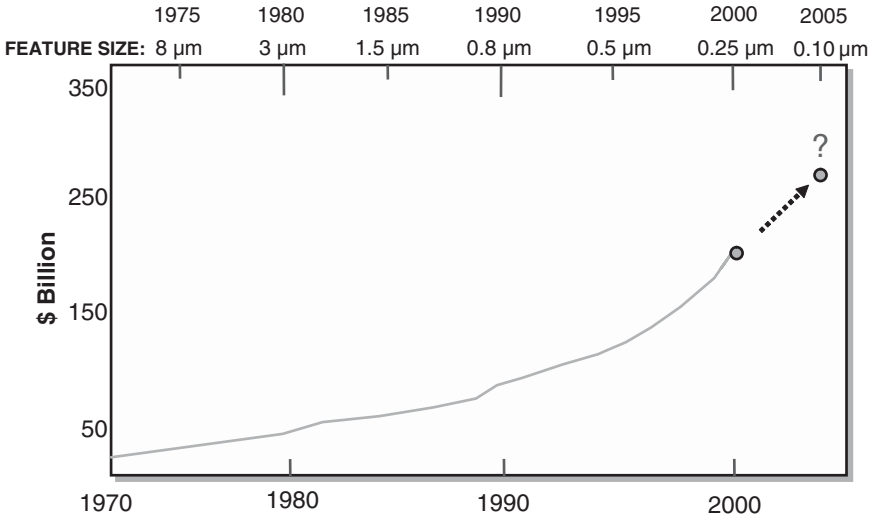


Figure 1-3: Semiconductor Industry Revenue with Chip Feature Sizes

Almost all of the devices made from semiconductors incorporate adjacent n-type and p-type regions in a variety of ways. The importance of putting these two oppositely doped regions right next to each other is discussed extensively throughout the book. In short, diodes, transistors of various types and other electrical components are made in this way.

1.2 The Semiconductor Industry

The semiconductor industry began in 1947 at Bell Laboratories in Murray Hill, NJ with the invention of the first transistor. At the time of this writing, the industry has grown to global proportions with annual chip sales exceeding $200 billion. The segment of the industry that makes processing equipment used to manufacture ICs is about $50 billion in annual sales.

Products include ICs, diodes, rectifiers and transistors made from semiconductor materials. ICs are now found in toasters, lamps, automobiles and most every electrical device. Of course, computers, office equipment, communications, defense, aerospace, industrial control and consumer electronics industries are dependent upon the semiconductor industry. Light emitting diodes (LEDs), lasers, photovoltaic (solar) cells, measurement and sensing devices are also part of the semiconductor industry.

SECTION 2

Support Technologies

The focus of this book is the story of how ICs are produced using the principal technology employed today, CMOS. However, the complexity of this technology requires the support of other equally sophisticated technologies. The story would not be complete without including a discussion of these fascinating topics.

2.1 Crystal Growth and Wafer Preparation

Single crystal silicon is the material most commonly used in the fabrication of integrated circuits. The silicon is purified from a surprisingly familiar starting material: sand. Many sand deposits are primarily silicon dioxide. The preferred types of sand contain very low levels of impurities. This starting material is further purified and chunks of very pure polycrystalline silicon are extracted.

Next, the silicon is melted in a large quartz glass crucible. The furnace must reach a temperature of at least 1414˚C (white heat), the melting point of silicon. A seed crystal in the form of a short rod about the size of a pencil with the desired atomic arrangement is lowered to touch the top of the melt. Silicon atoms attach to the seed crystal, slowly freeze and precisely replicate the atomic arrangement in the seed. The resulting ingot is gradually pulled up and out of the liquid melt. The ingot will often be 300 mm in diameter and one or two meters long.

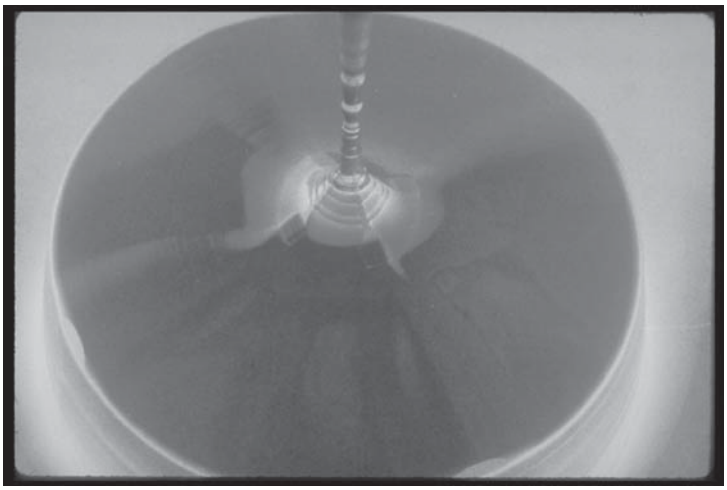


Figure 1-4: Silicon Ingot Being Pulled from Melt

The silicon ingot is machined on special tools that produce a more precisely shaped cylinder of the proper diameter. Then the ingot is cut into thin wafers using a diamond saw. The wafers are polished to a mirror finish using methods that minimize the damage done to the crystal structure at the surface. Most semiconductor devices use only a very shallow layer of the silicon at the wafer surface; maintaining good crystal properties is important to their performance.

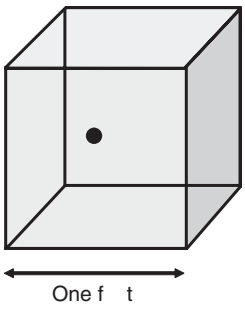
2.2 Contamination Control

Integrated circuits must be manufactured in a carefully controlled environment. The devices are so small that even a bacterium contaminating the surface of the chip would likely cause it to fail. Dust particles much smaller than the eye can see will produce "killer" defects. Particles can result in defects in the tiny patterns formed in building the device structures. Particles could also be a source of trace chemical contamination that may damage the product.

ICs are made in cleanrooms, often called the *fabrication area* or *fab* for short. Contamination levels are classified in terms of the number of particles found in a cubic foot of air that are 0.5 µm or more in diameter. The µm stands for micrometer, or micron for short, a unit of measure that is one one-thousandth of a millimeter in length. The Class 1 cleanroom, common to the industry today, will only have one of those particles in each cubic foot of air. Class 10 and Class 100 cleanrooms are also useful for less demanding technologies.

C eanroom C ass Number	# of Partic es > 0.5 µm/ft^3
1000	1000
100	100
10	10
1	1

Figure 1-5: Table of Cleanroom Classes



In a class-one cleanroom:
m re than ne particle
0.5 micr n r larger per cubic f t.

Figure 1-6: Class 1 Cleanroom

The cleanliness is maintained by the use of high efficiency particle attenuation (HEPA) filters. Ultra low particle attenuation (ULPA) filters are also used. These filters are positioned in the ceiling of the cleanroom or in the ceilings of special environmental chambers that must be kept contamination free. The air moves downward from the filters, producing a laminar flow pattern that will carry any stray particles downward through the return vents in or near the floor.

The cleanroom suit or "bunny suit" is critical to keeping the environment clean. It is a coverall made of nonparticulating material. The people working in the cleanroom are a major source of particles, so a complete coverall must be worn. Even makeup is forbidden because it produces a continuous cloud of particles emitted by the person wearing it.

The wafers upon which the chips are made are kept in closed containers during movement between manufacturing stations. The wafers are held in slotted cassettes, called boats, which keep them from touching each other. Automated robotic handlers are popular to move wafers through the fab. Each tool and inspection station is equipped with vacuum transfer arms or wands for handling individual wafers. The front of the wafer must never be touched except for processing!

2.3 Circuit Design and Mask Making

Chip design begins with a schematic diagram of the desired circuit. The schematic is translated into a picture or "layout" of the actual chip, drawn on a computer-aided design (CAD) tool. The chip is manufactured in layers; the overall layout is partitioned appropriately by the design software. Each layer is a pattern that must be reproduced on the wafer. These patterned layers, when stacked up, become the electronic components of the chip, all wired together.

The individual layer patterns are transferred to the chip using a photolithographic technique. A template, called a *photomask,* is made for each layer. The photomask is a picture or pattern etched in chrome on a quartz glass plate that allows light through in some areas and blocks the light elsewhere. Another form of the patterned plate is called a *reticle*. When light is projected through the photomask or reticle and onto a photosensitive material called *photoresist (resist)* coating the wafer, the pattern from the reticle will appear in the photosensitive material after treatment with a developer solution (see Section 3.2.3, Exposure, and Section 3.2.4, Develop). In effect, the photoresist-coated wafer acts like the film in a camera.

The reticles and photomasks are themselves made using a similar process. The mask-making company receives a computer file containing the design for each photomask or reticle. A blank, chrome-coated quartz plate is covered with photoresist. Then the plate is processed in a computer-controlled electron beam (e-beam) tool or a laser pattern generator that reproduces the desired pattern in the photoresist. Developer solution removes the unwanted resist. The remaining resist serves as the template (resist mask) for a wet etching process to remove the exposed chrome. When the pattern is reproduced in the chrome, the resist mask can be stripped off and the chrome mask (or reticle) is complete.

State-of-the-art chips require twenty to thirty such masks.

Photomasks are sometimes called *masks* for short so it can become confusing. Chapters 2 and 3 will clarify the terminology and elaborate on how the various masks are used.

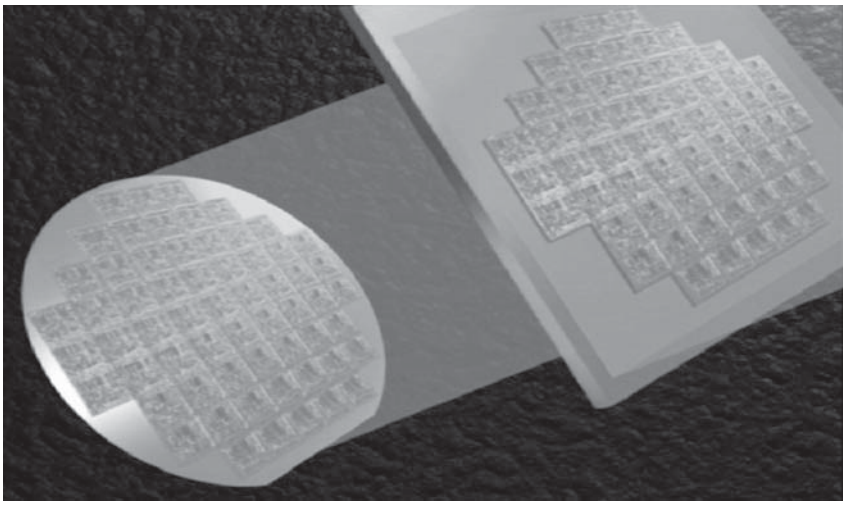


Figure 1-7: Exposing a Wafer Using Photomask

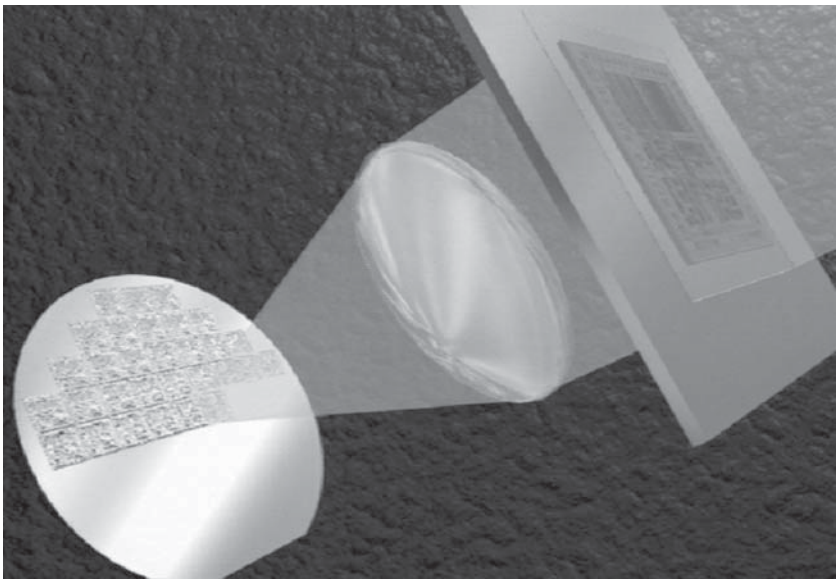


Figure 1-8: Exposing Wafer Using a Reticle

2.4 Process Diagnostics and Metrology

Strict dimensional control is critical to producing state-of-the-art ICs. Many of the devices available today to measure film thicknesses and line widths on the chips did not exist early in the history of the industry. It is no easy matter to measure these tiny dimensions. R&D work is underway for 32 nm technology, that is, a conducting "wire" on the chip that is 32 nm wide. The diameter of a hair is 50,000 to 100,000 nm! Chapter 3, Section 1 has the first of several helpful hints, found throughout the book, regarding the units of measure common to the industry. Here, nm stands for *nanometer* or one-billionth of a meter.

After each critical manufacturing step, the wafers are often inspected. Defects, contamination, processing or operator errors and metrology, and the measurement of critical dimensions (CDs) and film thicknesses, are some of the items that are checked at each inspection. Automated inspection stations do much of the work today.

The dimensions of features on today's chips are so small that even the sophisticated inspection and measurement tools developed in the last twenty years have been rendered at least partially useless. Scanning electron microscopes (SEM) are required to see the smallest features being produced today. Some films are so thin that even classical thin film measuring instruments like the ellipsometer can be inaccurate. Broad spectrum reflectrometry is still useful for film thicknesses, but sophisticated new technologies using ellipsometry are improving monitoring and measurements. New SEM technology is automated and makes the smallest measurements in the fab.

SECTION 3

Integrated Circuit Fabrication

The story of building IC's begins now.

3.1 Layering

Semiconductor devices are built in layers. While much of the critical electrical action occurs in the silicon wafer (substrate), a great deal of the chip building process involves the layering of thin films of various sorts on the substrate. Since chips are electrical devices, it is not surprising that layers of electrical conductors are separated from each other by layers of electrical insulators. There is a little more involved in building the whole device, as will be seen later, but that is the general theme of layering.

Thin film deposition is the general name for the technology of adding layers of various materials to the wafer. Layers are also formed by technologies called *oxidation* and *epitaxy*.

3.1.1 Insulators (Dielectrics)

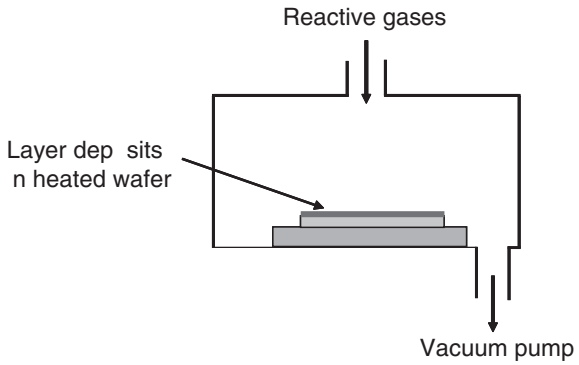
Silicon dioxide (oxide for short), the principle component of glass, is the most widely used insulator in the industry. Other chemicals are sometimes added to oxide to modify its physical characteristics.

Insulators are part of a more general class of materials called *dielectrics*. Dielectrics, as a class of materials, have a variety of properties that help make the chip work better. The text discusses several uses for these materials besides those of electrical importance.

To improve chip speed, special dielectrics are being developed to replace the oxide in some areas of the chip. These new materials are called *low-k* and *high-k* dielectrics. Their use is discussed later in the text.

Early in the wafer fabrication process, high temperature procedures can be used. Simply heating the silicon wafer in an oxygen atmosphere actually grows a coating of thermal oxide all over the wafer; some of the silicon is consumed by reacting with the oxygen to form silicon dioxide. Typically, a temperature of 800˚ to 1000˚C is needed for the oxidation reaction to proceed. The devices being made on the wafer cannot withstand that high of a temperature later on in the process, so lower temperature options must be used.

Chemical vapor deposition (CVD) is one of the processes of depositing a film on the wafer surface. The temperature can be much lower than that needed for thermal oxidation. As another example, some low-k dielectrics are applied in liquid form using a spinning chuck to hold the wafer; deposition is followed by a curing process at 300˚ to 400˚C. Other methods of low-temperature depositions are discussed in the text.



Schematic representati n.
There is a wide variety f chamber designs in use.

Figure 1-9: Chemical Vapor Deposition (CVD)

3.1.2 Semiconductors

Silicon technology is the topic of discussion for this book. Silicon is used in two forms: single crystal and polycrystalline (poly). The third form of silicon, amorphous silicon, is not typically used to make CMOS devices, and so it will not be discussed in this text.

The wafer is single crystal silicon. The production of the silicon ingot and the wafers was discussed in Section 2.1. Some chip designs require that the device be built in a thin layer of single crystal silicon that is grown on top of the starting wafer. That type of film is called *epitaxial silicon (epi)* and it is grown in an epitaxial reactor at a high temperature. The epi film is an extension of the underlying silicon crystal.

Silicon can also be deposited on the wafer, usually on top of a film of silicon dioxide. Polycrystalline silicon (poly) forms at temperatures above about 600ºC. It is composed of grains of single crystal silicon all stuck together. At lower deposition temperatures, the crystal regularity is lost and the silicon becomes amorphous.

Silicon in an IC almost always needs to conduct fairly well, so it will be doped to improve its conductivity. Doping is discussed in Section 3.3.

3.1.3 Conductors

Connecting the individual electrical components on the chip together to form a circuit is done mostly with metals, notably aluminum, copper and tungsten, but doped poly and silicides are used for some connections. Silicides are chemical compounds that can be made by reacting silicon and certain metals. These silicides conduct quite well and can withstand relatively high process temperatures.

Physical vapor deposition (PVD) is often the method for depositing metal on the wafer. The PVD tool uses large ions attracted to a target by an electric field. The target is made of the material to be deposited on the wafer. The bombarding ions eject atoms from the target. The atoms fly off the target (sputtering) and form a coating on the wafer. Of course, this “blanket” coating will have to be formed into conducting wires, if used for interconnections, as discussed in Section 3.2.

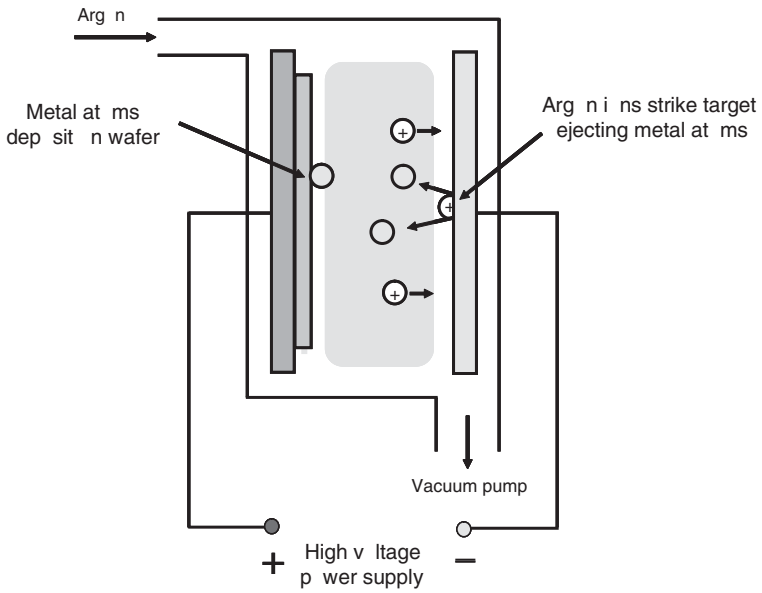


Figure 1-10: Physical Vapor Deposition (PVD)

A popular method of depositing copper is *electroplating*, also called electrochemical deposition (ECD). The wafer is immersed in a chemical solution containing copper sulphate. A copper plate is also immersed in the solution and when an electric current is passed through the solution between this plate and the wafer, copper is deposited on the wafer.

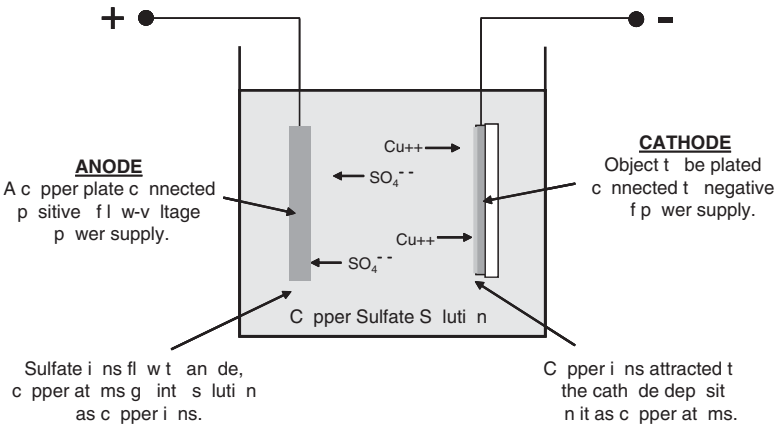


Figure 1-11: Electrochemical Deposition (ECD)

Because of the density of components on the IC, many crisscrossing layers of metal interconnects are needed. Each new layer is insulated from the previous underlying one by using a dielectric material. Holes are cut in the dielectric and filled with metal to make the electrical connections between layers. Eight layers of metal are common today and the number is increasing.

3.1.4 Chemical Mechanical Polishing

Chemical mechanical polishing (CMP) was mentioned earlier as the method used to polish newly produced starting wafers. It also has other uses in making ICs.

The optical tools discussed in Section 3.2.3 used to transfer the pattern from the reticle to the wafer are capable of printing very tiny features on the wafer. But the wafer becomes too bumpy rather early in the process for good pattern transfer; the optical tool cannot focus on all points on the wafer. The wafer must be planarized for best results. CMP is the best technology available for planarizing the surface of the wafer.

New processing methods and the use of copper have also required the use of CMP. The dual-damascene process discussed in the text requires that the copper be removed by polishing. The shallow trench isolation technology also must use CMP to remove excess oxide. Several examples will be discussed.

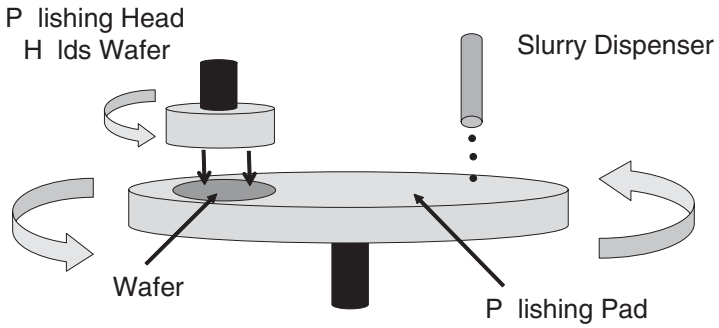


Figure 1-12: Chemical Mechanical Polishing (CMP)

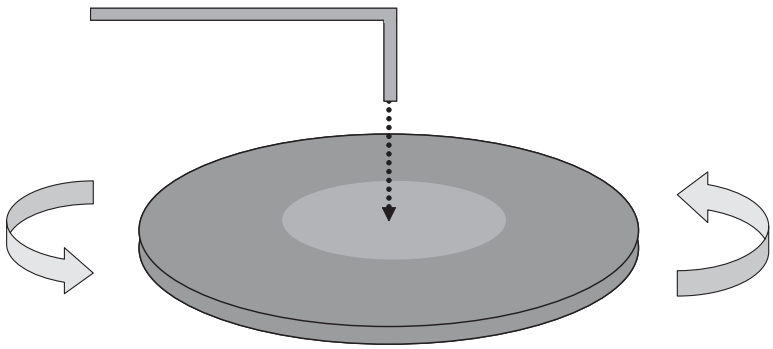
3.2 Patterning

The patterning process transfers the pattern from the reticle (or photomask) to the wafer using photolithographic technology. A photoresist template, also commonly called a mask, just to confuse everyone, is produced on the wafer surface that is a duplicate of the pattern on the reticle. One of two functions is performed by this template: it is used to protect parts of the wafer from chemical attack during an etching process or to block the implantation of dopants from the covered areas. The use of photolithographic techniques is also discussed in Section 2.3.

3.2.1 Photoresist Coat

Photoresist (resist) is a light-sensitive plastic material that is first dispensed in the form of a solution. The wafer is positioned on a holder or "chuck" that holds it in place with a vacuum while the chuck spins. The resist is applied to the wafer by dispensing a small amount of the thick, sticky liquid onto the center of the wafer. The spinning throws off excess resist and helps to determine the thickness of the resist film; a higher spin speed will drive off more resist and leave a thinner film behind. Other factors, such as the viscosity of the resist, contribute strongly to the film thickness.

A soft bake at about 100ºC drives off most of the solvent. The resist is stabilized (solidified) by the bake and the photosensitivity and solubility in developer solution are influenced, too.



Thin layer of photoresist solution spun on the wafer.

Figure 1-13: Photoresist Dispense

3.2.2 Exposure

Most exposure operations are done in a stepper. This tool holds the reticle and projects light through it. The resist is sensitive to ultraviolet (UV) light. The areas exposed to UV will dissolve in the developer solution. The unexposed areas are left behind and form a template or mask on the wafer.

The name "stepper" is short for the formal name of the technology: step-and-repeat reduction-projection printing. Step-and-scan is also a common technology found in the fab today. The image on the reticle is reduced in size by lenses. Most reduction systems in use today are 4×—that is, they reduce the image to 1/4 of its original size on the reticle.

The step-and-repeat or step-and-scan describes how the tool exposes the wafer. Since the image is reduced in size, only a small portion of the wafer can be exposed at a time. The wafer must be moved or stepped after each exposure. The wafer is exposed "step-wise" until the pattern has been transferred to the whole usable surface.

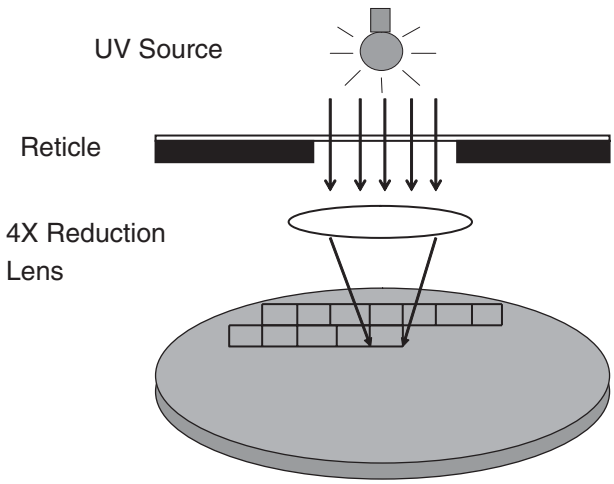


Figure 1-14: Photoresist Exposure Using Stepper

3.2.3 Develop

The exposed wafer is positioned on another spinning vacuum chuck, but this time a developer solution is sprayed on it. The exposed resist dissolves leaving the resist template (mask) behind.

3.2.4 Pattern Inspection

If an error in processing has created any imperfections in the photoresist image it can be stripped off and reworked. Now is the time to find any problems because the next processing steps will permanently transfer the pattern to the wafer. Imperfections could result in scrapping many wafers.

Automated inspection tools check the pattern for defects and measure important feature sizes called *critical dimensions* to monitor the quality of the product.

3.2.5 Etch

Most of the masking steps are followed by an etching step. Plasma etch is the dominant technology for permanently transferring the resist mask into the film or films below. The wafer enters an evacuated chamber. A gas is introduced and electrically charged (ionized) to change it into a plasma. The plasma is highly reactive and allows better control of the feature shapes and dimensions than the wet etch method.

Plasma Etchers

Technically, plasma etch reactors only make a small amount of the injected gas into a plasma. Typically, less than 1% of the gas particles are ionized at any one time. However, as discussed in Chapter 4, that is all that is needed to get the job done.

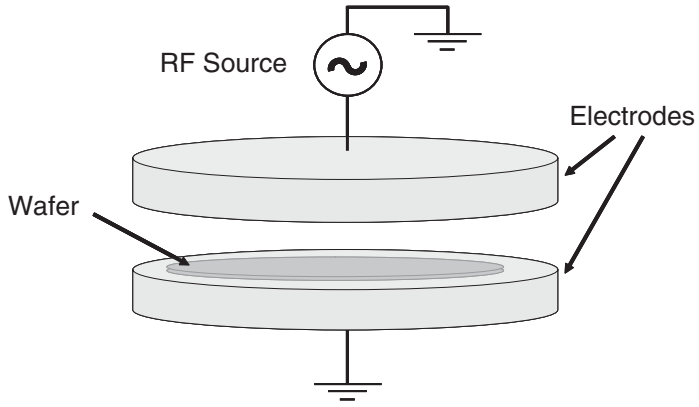


Figure 1-15: Plasma Etch

3.2.6 Implant

Some steps in the process use the resist mask as a protective template to permit dopant species to enter only selected areas. This implant mask is needed when an implanter is used to dope the wafer. An implanter is a particle beam machine that accelerates dopant ions and drives them into the wafer.

It is also common to etch the required implant masking template into a protective film (often a dielectric) that is deposited on top of the material to be doped. In this case, the process looks just like any other photo/etch step with the wafer going to implant after the etch sequence.

Section 3.3 will cover more on doping.

3.2.7 Photoresist Strip

When the resist mask is no longer needed it is removed, often by etching it off in an oxygen plasma (the resist is simply burned off) or through the use of a liquid stripper.

3.2.8 Etch Inspect

The features that were just produced on the wafer are inspected and critical dimensions measured. Defects found at this point often require the affected wafers to be scrapped.

The wafer is now ready to have another thin film deposited so it can return to patterning for the completion of another layer. The wafer loops through these steps until the device is complete.

3.3 Doping

Dopants are introduced to alter the electrical conductivity of the doped region. They can be added to the wafer in a variety of ways. The dopant is added to the silicon melt at crystal growth so the silicon ingot is either n- or p-type. The chemistry used to produce epitaxial silicon will usually include dopant species. The wafer can be doped simply by putting it in a furnace and introducing the dopant in gas form.

Today, almost all of the doping of state-of-the-art ICs is done by implanting the dopant using a particle-beam machine called an ion implanter. As mentioned above, the implanter accelerates ions down a long tube and drives them into the wafer. The

dopant enters the wafer wherever the implant mask has an opening. This method allows great precision in the placement of dopant as well as its concentration, or "dose." It is the doped regions that form critical sections of MOS transistors that will be discussed in detail in the text.

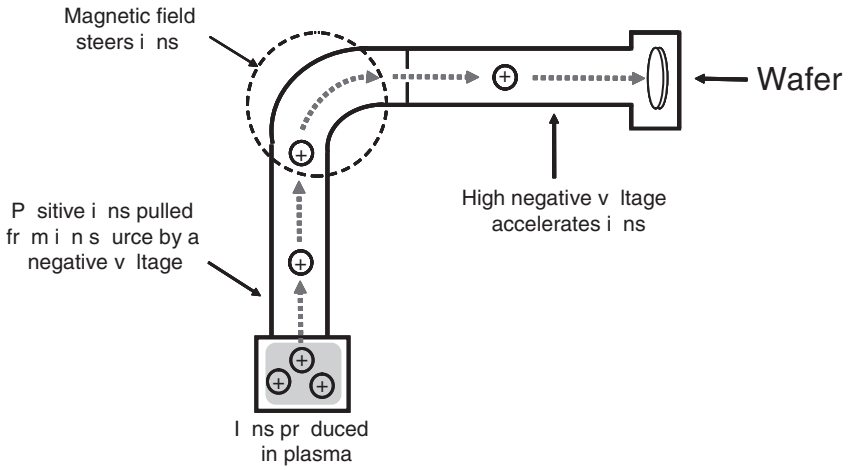
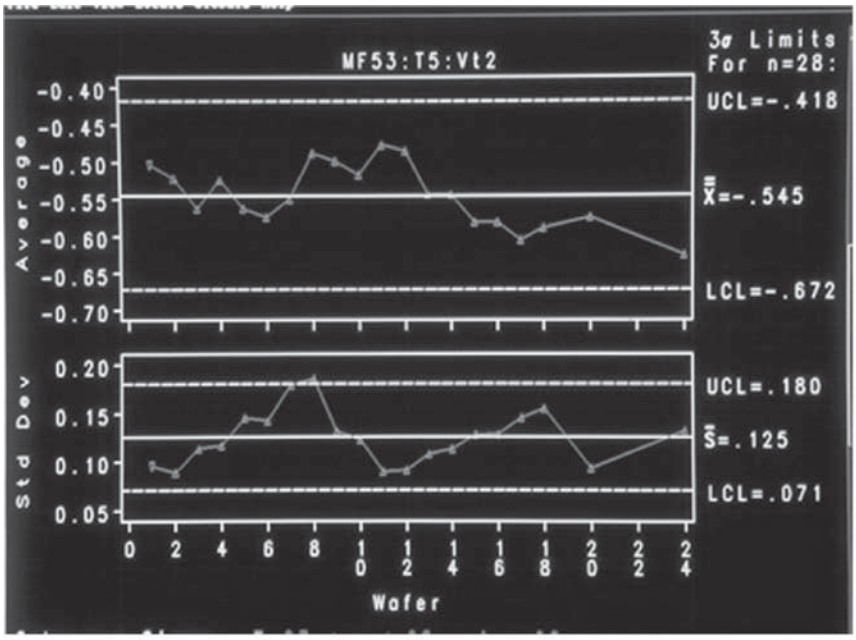


Figure 1-16: Ion Implantation Doping

3.4 Process Control and In-line Monitoring

Production lines must be constantly monitored in an attempt to prevent costly misprocessing. Statistical process control (SPC) is the best method for real-time monitoring of a manufacturing line. SPC charts are updated continually as each inspection is completed. The chart provides signals of a drifting process and often prevents expensive production errors.



KLA-Tencor

Figure 1-17: Statistical Process Control Chart

SECTION 4

Test and Assembly

The completed wafers are ready to become individual packaged parts if they pass electrical tests.

4.1 Electrical Tests

Two types of electrical tests are done while the wafers are still whole. First, several dozen basic measurements, called parametric tests, are made on individual transistors and other test structures in test patterns on the wafers. This is the first check on proper fabrication.

The second test is a circuit function test. Each chip is tested with a probe that has many tiny electrodes. A computerized system performs many tests on the chip and collects statistical data on the product. The test system keeps a map of passed and failed die on every wafer. Failed die are often marked with an ink dot or "inked out."

4.2 Die Separation

The chips are also called die (discussed in Chapter 8). The tested wafers are "diced" by putting each wafer on an adhesive holder and cutting the chips apart with a diamond saw.

4.3 Die Attach and Wire Bonding

The good die are removed from the adhesive and attached to a chip package. The dual inline package is the most common and contains a lead frame to which the chip is attached. The lead frame is a mounting that has the appropriate number and configuration of contacts or "leads" to attach to other electrical devices, circuit boards and the like.

The chip itself is attached to the lead frame using special solder such as gold-germanium alloy or epoxy. Then the tiny gold or aluminum wires that connect to the bonding pads on the chip are attached using special techniques that do not damage the pads.

4.4 Encapsulation

Encapsulation is the process of putting the chips into protective packages. Many packages are available for chips. The intended use of the part is key in deciding how to encapsulate it. Military hardware is exposed to harsh environments so tough packaging is used. Simple plastic packages, such as the dual inline package, suffice for many chips. Metal, ceramic and composite packaging material are all used.

4.5 Final Test

The packaged parts receive a functional test similar to that given earlier to the die on the wafers, but it is usually more extensive since many products cannot be tested at full speed or power until final assembly is complete.

The parts are now ready for shipping to stores or customers.

SECTION 5

Summary

Chapter 1 summarized the components of IC production. Chapter 2 begins the story of building a chip from the ground up. In Chapters 3 through 7, the reader will actually build a chip, figuratively speaking, taking the wafers through the entire process, watching each component of the chip come together.