CHAPTER 2

TECHNOLOGY

2.0 INTRODUCTION

A good understanding of processing and fabrication technology on the part of the circuit designer is necessary to provide the flexibility needed to optimize integrated circuit designs. With this knowledge the actual layout can be considered during design and the appropriate parasitics can be included in the analysis. Innovative techniques that improve performance often involve circuits or geometries that are dependent on and applicable to a particular process. Knowledge of processing characteristics enables the designer to make yield calculations during design and consider tradeoffs between yield, performance and design simplicity.

In this chapter processing technology is discussed from a qualitative viewpoint. This is followed by a detailed discussion of typical NMOS, CMOS, bipolar, thick film, and thin film processes. Most processes in industry can be viewed as either a straightforward variant or extension of these processes. These processes are summarized in the appendices of this chapter. Included in the appendices are process scenarios, graphical process descriptions, design rules, process parameters, and some computer simulation model parameters. These appendices should provide a useful reference for material that is presented in later chapters of this book. This chapter is concluded with a discussion of practical layout considerations and comments about some CAD tools that have become an integral part of the IC design process.

2.1 IC PRODUCTION PROCESS

The major steps involved in producing integrated circuits are considered from a qualitative viewpoint in this section. These steps are used in the MOS and/or bipolar processes that will be discussed later in this chapter.

2.1.1 Processing Steps

CRYSTAL PREPARATION. The substrate of bipolar and MOS integrated circuits is generally a single crystal of silicon that is lightly doped with either n- or p-type impurities. The substrate serves both as the physical medium upon and within which the IC is built and as part of the electrical circuit itself. These crystals are sliced from large right-circular cylinders of crystalline silicon, which are carefully grown to lengths up to 2 m and which vary in diameter from 1 to several inches. The slices are typically 250 μ to 400 μ thick. From an electrical viewpoint much thinner slices would be acceptable; however, the thicker slices have been adopted because they are more practical to handle (less breakage) and are less likely to warp during processing. The size of the wafers has been increasing rapidly with time to allow for both large chips and a larger number of chips per wafer. As of 1989, many of the older processing lines were using 4 inch wafers, but the newer lines are typically using 5 and 6 inch wafers. The crystals are often cut so that the surface is oriented approximately in the ‹100› direction.

MASKING. IC masks are high-contrast (black on clear) photographic positives or negatives. They are used to selectively prevent light from striking a photosensitized wafer during the photolithographic process. The masks are typically made of glass covered with a thin film of opaque metal, although less costly and less durable emulsion masks are sometimes used. The masks are produced from a digitized description of the desired mask geometries. There are several different methods of generating the masks (called pattern generation) from the digitized circuit description. One method involves photographically reducing large copies of the desired patterns that have been generated with a computer-controlled drafting machine. This method was used widely in the past but has largely been replaced by the next two. A second uses a laser beam as a pattern generator in a raster-scan mode. Both of these methods generally also require a high-resolution step and repeat and/or reduction camera to make the final masks that will be used. The intermediate image that is created is called a reticle and is usually 5 or 10 times real size. A third method uses an electron beam (E-beam) to generate the actual patterns directly onto the final masks. This method produces the best quality masks and is used extensively for very small geometries, but it requires considerable time and expensive equipment.

PHOTOLITHOGRAPHIC PROCESS. Photoresist is a viscous liquid. It is applied in a thin, uniform layer (about 1 μ thick by spinning the wafer) to the entire surface of a wafer following cleaning. After application the photoresist is hardened by baking. The physical characteristics of the photoresist can be changed by exposure to light. The photoresist thus acts as a film emulsion and can be exposed by light through the transparent areas of a mask (either by contact printing or projection), by a projection of light through a reticle containing the same information (called direct step on wafer), or by an electron beam (E-beam) that scans the desired regions. Following exposure, the resist is developed to selectively remove the resist from unwanted areas. This step is often followed by another baking to further harden the remaining photoresist.

Both positive and negative photoresists are available. With negative photoresist the unexposed areas are removed during development, and with positive resist the exposed areas are removed. Negative resists are noted for being quite unaffected by etchants used in processing, but finer resolution can typically be obtained with positive resists. Photoresists serve as protective layers to many etchants and oxidizing agents, and as a barrier to ion implants.

Proper mask alignment is essential to maintain device operation, characteristics, and yield. Alignment markings are generally included with the circuit information when the masks are made so that these marks will appear on the wafer during and after processing. A machine called a mask aligner is used to align and expose the wafers. Figure 2.1-1 shows typical alignment characteristics. The physical size and geometry of the masks used for fabrication is governed by the particular technique used by the mask aligner to expose the wafer. Mask aligners that use contact printing have multiple copies of the individual circuits at actual size (i.e., 1×) accurately patterned on the mask. These aligners have a large throughput and are relatively inexpensive. The large masks, however, have a very short lifetime (typically 3 to 10 exposures) because of damage incurred when the mask contacts the photoresist for exposure. This increases effective mask costs. The direct step on wafer aligners typically use a 5× mask (often called a reticle) as a negative. It typically will contain only a single copy of the circuit, though several copies may be used for small ICs. The image is optically reduced to 1× upon exposure. The wafer must be repeatedly moved to the next location after each exposure until the entire wafer is exposed. The lifetime of the mask is very long since no physical contact is made, but the throughput has been decreased considerably to allow for the successive wafer movements. The

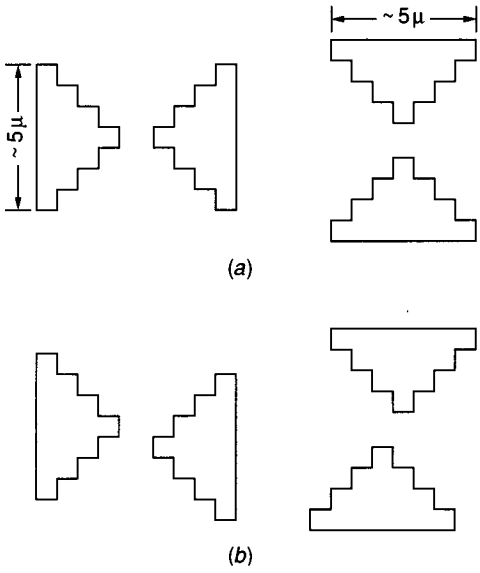

FIGURE 2.1-1
Alignment marks: (*a*) Mask marks, (*b*) Simulated positioning of alignment marks after fabrication.

equipment is also considerably more expensive because of the precision needed to maintain consistent and repeated uniform stepping of the wafer. Both types of aligners are widely used in industry.

One of the most practical and popular methods of exposure actually combines the mechanical economics of the 1× aligners and the mask life of the steppers. In this approach, a thin protective membrane, called a pelicle, is placed above the emulsion of 1× chrome masks for protection of the mask and long mask life. Although the membrane itself may get dirty or scratched, it is placed far enough away from the mask so as to remain out of focus and thus not project defects onto the wafer when columniated light is focused through the mask onto the wafer.

A fourth method of exposure actually uses no masks at all. Instead, a narrow electron beam (E-beam) is selectively focused on the wafer in a raster-scan manner in small regions, with wafer stepping to position successive portions of the wafer under the beam. The same digital database that is used to generate masks can be used to drive the E-beam system. This approach gives better resolution than any of the previously discussed methods but involves very expensive equipment and has a much smaller throughput. It is practical for only the most demanding applications.

DEPOSITION. Films of various materials must be applied to the wafer during processing for most existing semiconductor processes. Often these films are very thin (200 Å or less for some SiO_2 layers) but may be as thick as 20 μ for "thick film" circuits. Films that are deposited include insulators, resistive films, conductive films, dielectrics, n- and p-type semiconductor materials, and dopants that are subsequently forced deeper into the substrate. Deposition techniques include physical vapor deposition (evaporation and sputtering), chemical vapor deposition (CVD), and screen printing for the thick films. With the exception of the screen-printed films, the depositions are nonselective and are placed uniformly over the entire wafer.

Evaporation refers to evaporating the material that is to be deposited by controlling the temperature and pressure of the host material environment. A film is formed when the material condenses. A continuous evaporation–condensation process is established that allows for a controlled growth rate of the film.

Sputtering involves bombardment of the host material with high energy ions to dislodge molecules, which will reattach themselves to the surface of the wafer (as well as to other surfaces in the sputtering apparatus). Often two different host materials are simultaneously bombarded at different rates to establish the characteristics of the sputtered material. This dual host bombardment is termed cosputtering. With some materials, sputtering offers advantages over evaporation in host material integrity on the deposition surface.

Chemical vapor deposition (CVD) is achieved in two ways: (1) by causing a reaction of two gases near the substrate, a reaction occurs that creates solid molecules, which subsequently adhere to the substrate surface; or (2) by pyrolytic decomposition (a decomposition caused by heating) of a single gas, which also frees the desired molecules for reattachment.

ETCHING. Etching refers to selectively removing unwanted material from the surface of the substrate. Photoresist and masks are used to selectively pattern (expose) the surface of the substrate. Following this patterning, the physical characteristics of the surface are changed by etching. A single IC will generally undergo several different etches during processing. The chemicals used for etching are chosen to selectively react with unprotected areas on the wafer while not affecting the protected areas. A summary of the effects of some commonly used etchants on typical semiconductor materials is shown in Table 2.1-1.

There are two types of etches used in production: wet and dry. The *wet etches*, often called chemical etches, use liquid etching agents, which are applied to the substrate surface. Although they have received widespread application in the past, they etch horizontally as well as vertically into the surface of the substrate. This horizontal etching causes undercutting of the patterned areas. Unless the width of the nonetched regions is orders of magnitude greater than the thickness of the material being etched, the nonuniformity of the horizontal etching causes significant changes in desired device characteristics.

TABLE 2.1–1
Characteristics of commonly used fabrication materials

I. Materials used in IC fabrication		
Purpose	**Materials**	**Comments**
Silicon crystal substrates	$SiCl_4$	Silicon source for growth of single crystal silicon
	$SiHCl_4$	Silicon source for growth of single crystal silicon
	SiO_2 (Sand)	Silicon source for growth of single crystal silicon
Silicon layers (both single crystalline and polysilicon)	$SiCl_4$ and H_2	The hydrogen gas strips the Cl atoms to form solid silicon.
	SiH_4	Heat causes the release (pyrolysis) of H_2 gas.
	SiH_2Cl_2	Heat causes the release (pyrolysis) of HCl gas.
Oxides	O_2	Used to grow SiO_2 by thermal oxidation
	H_2O (Steam)	Used to grow SiO_2 by thermal oxidation
	SiH_4 and O_2	Used for CVD deposition of SiO_2 and to grow protective "glass" (SiO_2)
Nitride layers	Si_4 and NH_3	The ammonia causes the release of hydrogen gas and leaves Si_3N_4.
	$SiCl_4$ and NH_3	The ammonia causes the release of HCl and leaves Si_3N_4.
Etches, wet	HF	Hydrofluoric acid etches SiO_2 but not Si, Si_3N_4, or photoresist.
	HF and HNO_3	Etches Si

TABLE 2.1–1
(Continued)

Purpose	Materials	Comments
	H_3PO_4	Hot phosphoric acid etches Si_3N_4 but not SiO_2. Removes some types of photoresist.
Etches, dry	CHF_3	Etches SiO_2
	C_3F_8	Etches SiO_2
	SF_6	Etches silicon
	CF_4	Etches Si_3N_4
	CCl_4	Etches aluminum
Patterning	Photoresist	Used as barrier to ion implants. Also used to pattern SiO_2 since photoresist is not affected by HF, a common SiO_2 etchant.
	SiO_2	Acts as a barrier to some p- and n-type impurities
	Si_3N_4	Used as protective layer over silicon or SiO_2 to prevent thermal growth of SiO_2. Also serves as a barrier to low-energy ion implants although thin layers can be and are penetrated with higher-energy implants. Also serves as a diffusion barrier to impurities such as Ga, Al, Zn, and Na.

II. Sources of impurities

	Impurities	Source
n-type	Arsenic	As_2O_3, AsH_3
	Antimony	Sb_2O_3, Sb_2O_4
	Phosphorus	P_2O_5, $POCl_3$ (liquid), PH_3 (gas—implant or diffusion)
p-type	Gallium	
	Aluminum	
	Boron	BN (solid), BBr_3 (liquid), B_2O_3 (gas), BCl_3, (gas), B_2H_6 (gas), BF_3 (for implants)

III. Impurity migration in silicon

Impurity	Silicon	SiO_2
Arsenic†(n)	moderate	very slow
Antimony (n)	moderate	very slow
Phosphorus (n)	fast	slow
Gallium (p)	moderate	fast
Aluminum (p)	fast	fast
Boron (p)	fast	slow

†Arsenic is often preferred to the other n-type impurities because it gives more abrupt junction gradients, which yield better frequency response and improved current gain in bipolar transistors. Due to environmental concerns, however, the use of arsenic in the semiconductor industry is limited.

Dry etching, also termed ion etching, is directional and thus much less susceptible to the undesirable horizontal undercutting. Dry etching techniques include sputter etching, ion-beam etching, and plasma etching. Since no liquid chemicals are involved, a significant reduction of costs associated with disposal of spent chemicals is realized when dry etches are used. The etch rate for dry etches is generally lower than the wet etch rate. Dry etching is recognized as a practical alternative to wet etching and is widely used.

The characteristics of an ideal wet etch and an ideal directional etch are shown in Fig. 2.1-2. The nondirectional etch is termed an isotropic etch. The edge profile appears approximately circular, with radius r and center at point A. If the etch is stopped precisely when the underlying layer is exposed the radius r will be T, the thickness of the layer, and the undercut of the protective layer, X, will be also T. If the etch is not stopped precisely when the underlying layer is exposed, the radius will be T_1, which is greater than T, and both the effective opening and the undercut will thus be larger than desired.

An ideal directional etch is termed an anisotropic etch. Note that an anisotropic etch has a very abrupt edge, which causes problems for applying subsequent layers uniformly and reliably across this edge.

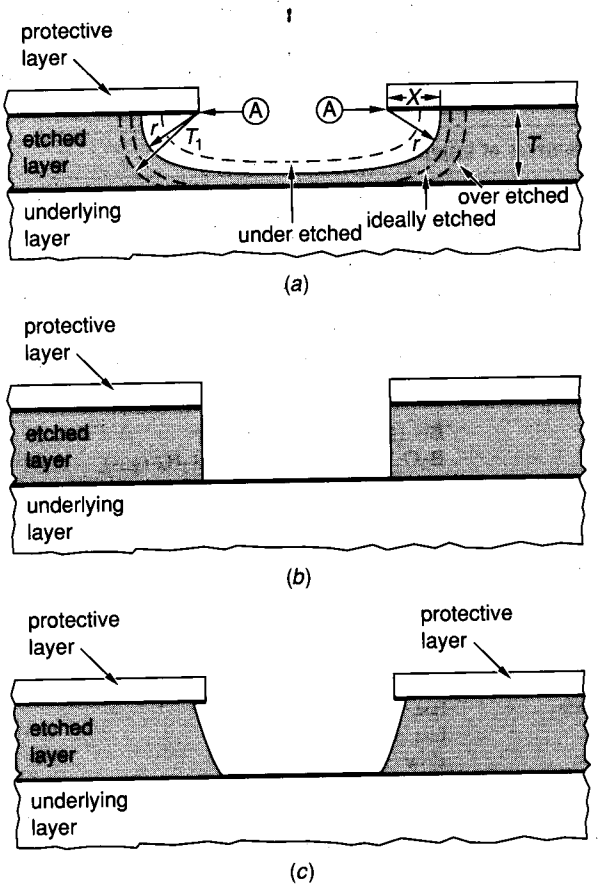

FIGURE 2.1-2
Characteristics of etches: (*a*) Isotropic etch, (*b*) Anisotropic etch, (*c*) Preferential etch.

The term preferential etch characterizes an etch that prefers one direction but is less directional than an anisotropic etch. A preferential etch is depicted in Fig. 2.1-2*c*. The preferential etch may be preferable where subsequent layer coverage of an anisotropic etch is problematic.[1]

DIFFUSION. Diffusion, in the sense of an IC processing step, refers to the controlled forced migration of impurities into the substrate or adjacent material. The resultant impurity profile, which plays a major role in the performance of the integrated circuit, is affected by temperature and time as well as the temperature–time relationship during processing. Subsequent diffusions generally cause some additional migration of earlier diffusions. Actually, the diffusion process continues indefinitely, but at normal operating temperatures of the integrated circuit it takes tens of years or longer for the additional movement to become significant.

The method by which the impurities are introduced varies. A solid deposition layer or a gaseous layer above the surface can be used as the source of impurities. Impurities can also be accelerated to selectively bombard the substrate so that they actually become lodged inside the substrate very near the surface. This technique, termed *ion implantation*, offers very accurate control of impurity concentrations but causes significant crystal damage near the surface.

The purpose of a diffusion following deposition is to cause a migration of carriers into the substrate from either solid or gaseous surface layers. A diffusion step following ion implantation is used to mend or anneal bombardment–induced fractures in the single crystalline structure at the surface of the substrate as well as to cause additional impurity migration.

As in the etching process, the direction of impurity diffusion is difficult to control with accuracy. Impurities typically diffuse both vertically and laterally from the surface at comparable rates in a manner similar to that observed for the isotropic etch of Fig. 2.1-2*a*.

CONDUCTORS AND RESISTORS. Aluminum or other metals are often used as conductors for interconnection of components on an integrated circuit. These metals are typically deposited, patterned, and etched to leave interconnects where desired. The thickness of the popular aluminum films is typically about 6000–8000 Å but may be as much as 20,000 Å for linear (analog) single-level metal processes. Metal films are particularly useful for interconnects that must carry large currents, but traces must be wide enough to avoid the *metal migration*, or *electromigration* problem. Electromigration is the movement of atoms with current flow and can be likened to wind erosion of dirt. If significant metal migration occurs, the conductors become open, resulting in failure of an integrated circuit. Metal migration is insignificant provided the peak current density in the conductor is below a certain threshold. For aluminum, this threshold is around 1 mA/μ^2. This threshold is material dependent and ranges from 0.05 mA/μ^2 to 2 mA/μ^2 for other similar materials.

Nonmetallic films are widely used for conductors and interconnects when current flow is small. These materials are typically worse conductors than metals

and thus cause a significant voltage drop when currents are large. These materials find limited applications as resistors.

Polysilicon is one of the most popular nonmetallic conductors. Polysilicon differs from single crystalline silicon, which is often used as a substrate material, only in that polysilicon is composed of a large number of nonaligned, randomly oriented, small silicon crystals. Although polysilicon is chemically identical to single crystal silicon, its electrical characteristics are much different. Polysilicon is a good conductor when heavily doped and a good resistor when lightly doped. Polysilicon is often used for gates of field effect transistors (MOSFETs) and as an electrode for capacitors. Polysilicon can be deposited over SiO_2. SiO_2 can also be readily grown on polysilicon and is often used to serve as a dielectric and isolate two polysilicon layers in processes where double polysilicon (*double poly*) layers are available. Polysilicon's characteristics are dependent on the size of the small crystals, often termed the grain size. It can be deposited on a variety of materials and the growth rate can be fairly fast. Polysilicon films are typically about 2000 Å thick and are often termed *poly*.

Silicides and/or refractory metals are often used on top of or in place of polysilicon for fabricating conductors. These materials are often much better conductors than polysilicon.

OXIDATION. Oxidation is the process whereby oxygen molecules from a gas above the substrate or surface material cause the growth of an oxide on the surface. Since the substrate or surface material is typically silicon, the oxidation process produces silicon dioxide. The speed at which the SiO_2 layer grows is a function of the doping concentration and the temperature of the substrate during oxidation. The SiO_2 layer serves as a very good insulator between the substrate or surface material and whatever is placed upon it. When the SiO_2 layer is grown on the substrate, a small amount of the Si in the substrate is consumed to provide for the Si molecules in the oxide. The growth of x microns of SiO_2 consumes approximately $0.47x$ microns of single crystal silicon.

As an alternative to oxidation, the SiO_2 layer can be applied by CVD. This technique is used extensively when the SiO_2 layer must cover something other than Si since no silicon molecules are available for oxidation. CVD can also be done at lower temperatures, which is advantageous if additional diffusion of previously deposited materials must be minimized. SiO_2 layers formed by oxidation are generally more uniform than those formed by a CVD process.

Other types of oxides are also used as insulating layers in fabricating ICs. Doped deposited oxides such as phosphosilicate glass (PSG) are often used as insulators on top of polysilicon. Some of these are doped to improve reflow characteristics during annealing. This doping helps reduce sharp boundaries (improve step coverage) introduced during etching of polysilicon.

Nitride (Si_3N_4) is also used as a dielectric between two levels of polysilicon in some processes. The dielectric constant of Si_3N_4 is about four times that of SiO_2. This offers potential for much higher capacitance densities for fixed dielectric thicknesses or much thinner dielectrics for a fixed capacitance density. A thin layer of SiO_2 is generally applied to the Si prior to the Si_3N_4 to minimize the mechanical stress associated with a direct Si interface to Si_3N_4. This stress

is caused by a difference in the lattice characteristics of single-crystal silicon and the Si_3N_4 layer.

Although somewhat different chemically, polyimides are also used as insulating layers, most notably between two metal layers. Polyimides tend to smooth abrupt underlying irregularities, thus reducing the effects of sharp boundaries of underlying metal layers.[2]

EPITAXY. Epitaxial growth is generally a CVD. It warrants singling out, however, because of the extensive use of this process step in bipolar integrated circuitry and because epitaxial layers are ideally single crystalline extensions of the substrate. Epitaxial layers are grown slowly enough that the molecules added to the surface can align with the underlying crystalline structure of the substrate to form a crystalline epitaxial layer. A small amount of n- or p-type impurities is generally intentionally introduced into the epitaxial layer during the epitaxial growth to obtain a doped epitaxial layer.

Several excellent textbooks provide considerably more detail about semiconductor processing; the interested reader may consult References 3–6.

2.1.2 Packaging and Testing

After processing, the integrated circuits are tested and packaged. The first step in the testing process generally involves a process verification to make certain that the process parameters are within the tolerances acceptable for the product. To facilitate this verification, *test plugs* containing special test structures specially designed for this purpose are included on the wafer at several locations in place of the regular circuits themselves. Alternatively, to avoid sacrificing the potential production die sites that are devoted to test plugs, there is a growing trend to integrate the test patterns into the *scribe lines*, which are existing grid lines void of circuitry where cuts will be made to separate the dies. A wafer prober is used to make mechanical contact with the test plugs so that electrical measurements can be made. Assuming the process parameters are within tolerance, the individual dies are automatically probed and electrically tested. Defective dies are marked with ink and later discarded. After probing, the wafer is scribed (typically with a wafer saw) both horizontally and vertically between adjacent dies, and the dies are separated. Following separation the individual dies are *die attached*, or *die bonded*, to a carrier or to the IC package itself. Wire bonds are subsequently made from the pins of the package to the appropriate locations on the die. The bonding wires are typically of either gold or aluminum. The diameter of this wire is in the range of 1 mil. After the wire bonds are complete, the packages are formed or closed and a final electrical test (and burn in for some parts) is completed.

Packaging technology saw minimal advancements through the late 1970s and early 1980s. It is well recognized that existing packaging techniques are a major bottleneck in the evolution of IC technology. Considerable effort on a worldwide basis is focused on the packaging problem. Practical alternatives to the conventional packaging approach, described above, will likely evolve in the next few years.

2.2 SEMICONDUCTOR PROCESSES

There are currently three basic processes used for the fabrication of monolithic integrated circuits containing active devices. These are the NMOS, CMOS, and bipolar processes. The first two are both termed MOS (Metal Oxide Semiconductor) processes even though, as will be discussed later, the standard acronym is no longer completely descriptive. A fourth approach essentially combines bipolar and MOS technologies into a single but more involved process. The mixed bipolar–MOS process is called Bi–MOS. A fifth method for constructing ICs is termed the hybrid process. These processes are depicted in Fig. 1.2-1.

Generic processes similar to those used in industry will be discussed in this section. The generic NMOS and CMOS processes discussed are very similar to those available through MOSIS† and the same terminology and conventions that have been established by MOSIS will be followed when practical. Several excellent references provide additional information about the NMOS and CMOS processes available through MOSIS and about MOS processing in general.[1, 2, 7–11] Additional information about these generic processes, such as design rules and process parameters, are discussed in Section 2.3. Details about a typical Bi–MOS process are not presented, but the basic approach should be apparent after studying the basic MOS and bipolar processes.

The NMOS (n-channel MOS) and CMOS (Complementary Metal Oxide Semiconductor) processes are quite similar in that both have the field effect transistor (FET or MOSFET) as the basic active device. In the NMOS process n-channel MOSFETs are available as the active devices whereas in the CMOS process both n-channel and p-channel devices are available. When compared to the NMOS process, the CMOS process offers advantages in design simplicity at the expense of more processing steps. It is often the case that CMOS also offers improvements in power dissipation and performance and in some cases even size over NMOS. These tradeoffs must be considered when selecting the most economical process for a given application. Another MOS process, PMOS, is available but will not be singled out because it is essentially a dual of the NMOS process. In the PMOS process the basic active device is the p-channel MOSFET. Although the PMOS process was commonly used for some of the earlier MOS circuits, the NMOS process offers some advantages due to characteristics of

†The MOSIS (MOS Implementation System) program is sponsored by the U.S. Department of Defense Advanced Research Projects Agency (DARPA). Academic participation in this program has been supported by NSF since 1981. Authorized participants undertake designs in one of several MOS processes supported by the program and submit these designs to MOSIS for fabrication. Designs from a large number of different institutions are combined on a multiproject chip format to significantly reduce the fabrication costs below what would be experienced if designs were independently fabricated. MOSIS assumes responsibility for both mask generation and processing. Eight to ten weeks after the published closing date for a specific processing lot, the designer is scheduled to receive packaged parts and occasionally some unpackaged dies. Technical information about these parts is also included. Additional information about the MOSIS program can be obtained from USC Information Sciences Institute, 4676 Admiralty Way, Marina Del Rey, CA 90292–6695, (213) 822–1511.

semiconductor materials available (specifically, electron mobility is higher than hole mobility) and is more popular today.

The bipolar process is so named because the basic active device is the Bipolar Junction Transistor (BJT). Higher speeds are currently available with the bipolar process than for the NMOS and CMOS processes although significant improvements in the speed of the latter processes have been and continue to be made. Bipolar integrated circuits are noted for their considerable internal power dissipation compared to that of the NMOS and CMOS processes. For logic circuits the NMOS and CMOS circuits have a significantly higher component density than their bipolar counterparts.

The hybrid process combines thin and/or thick film passive components that are on one or more separate substrates with active devices from a separate substrate onto a common carrier. This makes hybrid ICs quite expensive. For applications that require precise and temperature-stable passive components, the hybrid process often offers a practical solution.

MATERIAL CHARACTERIZATION. Some terminology that is common to most semiconductor processes is best introduced at this point. Throughout this text the notation n^+ will denote a heavily doped n-type semiconductor region, and n^- will denote a lightly doped region. The designation n^+ or n^- will be assumed relative to the context in which this designation is made. No superscript will be included if the region is doped somewhere between n^+ and n^- or if it is not necessary to make the distinction in the given context. The same convention will be followed for p, p^+, and p^- designations.

The *resistivity* of a homogeneous material is a volumetric measure of resistive characteristics of the material. The resistivity is typically specified in terms of ohms-cm (Ω-cm). If a right rectangular solid of material of length L and cross sectional area A (see Fig. 2.2-1*a*) has a measured resistance of R between the two ends, then the resistivity of the material is given by

$$\rho = \frac{AR}{L} \tag{2.2-1}$$

where it is assumed that the contacts on the two ends cover the entire surface and are perfectly conducting.

The *sheet resistance* is a measure of the characteristics of a large, uniform sheet or film of material that is arbitrarily thin. The sheet resistance is specified in terms of *ohms per square* of surface area. If a rectangular sheet of material of length L and width W (see Fig. 2.2-1*b*) has a measured resistance R between the two opposite ends, then the sheet resistance of the material is given by

$$R_\square = R\frac{W}{L} \tag{2.2-2}$$

where it is assumed that the contacts on the two edges cover the entire edge and are perfectly conducting. It should be emphasized that both the resistivity and sheet resistance are characteristics of materials independent of particular values for A, L, W, and R in the previous equations.

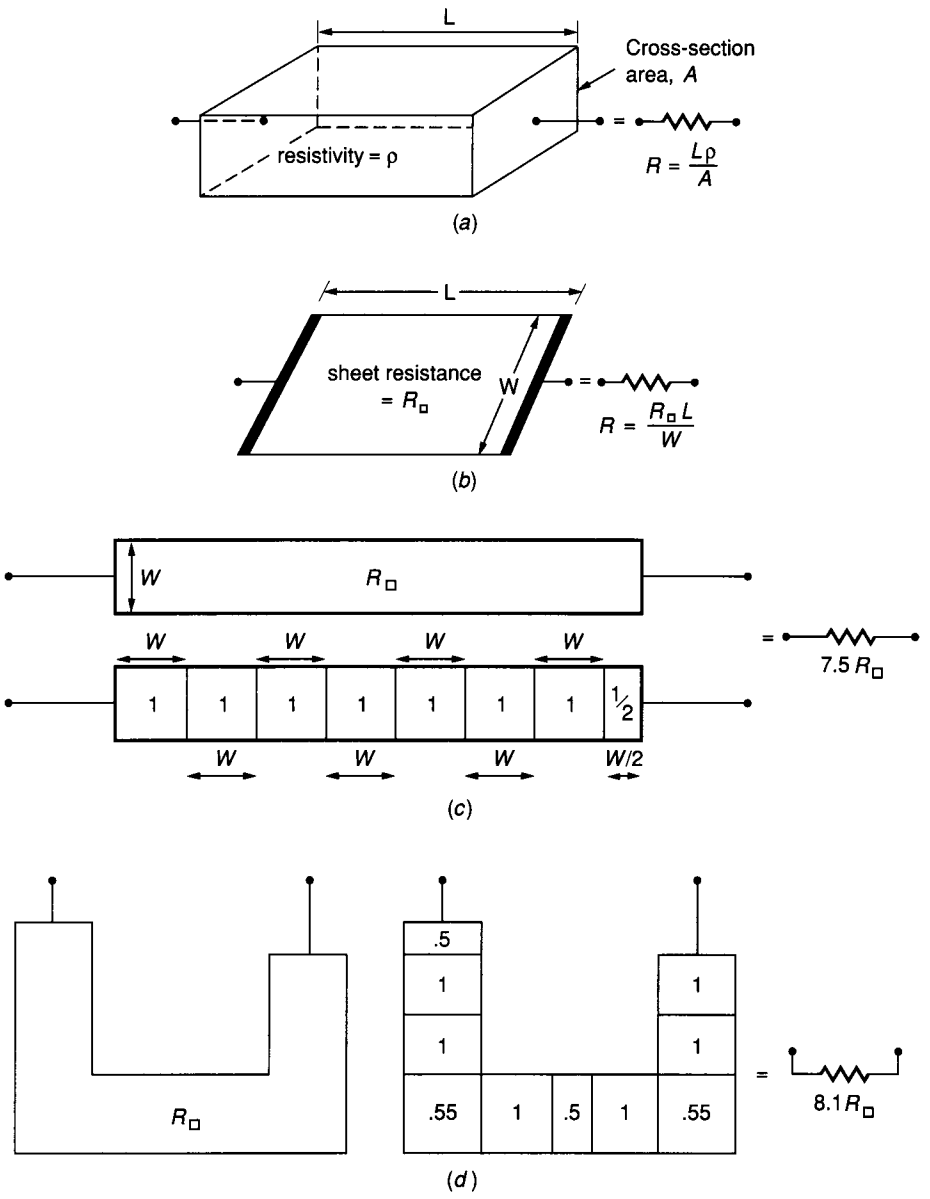

FIGURE 2.2-1
Resistive characteristics of bulk and sheet materials: (*a*) Resistivity, (*b*) Sheet resistance, (*c*)&(*d*) Graphical calculations from sheet resistance.

Example 2.2-1. Determine the length of a 100 kΩ rectangular resistor that is 25 μ wide and is to be constructed from a thin sheet of material with a sheet resistance of $R_\square = 100\Omega/\square$.

Solution. From Eq. 2.2-2,

$$L = \frac{WR}{R_\square} = (25\ \mu)\left(\frac{100\ \text{k}\Omega}{100\ \Omega}\right) = 2.5\ \text{cm}$$

Note the excessive length of the 100 kΩ resistor!

The sheet resistance of a thin layer of thickness z constructed from a material that has a resistivity ρ is given by

$$R_\square = \frac{\rho}{z} \tag{2.2-3}$$

The resistance of thin rectangular regions of length L and width W on an integrated circuit can be readily obtained from the sheet resistance by counting the number of square blocks of length W that can be placed in the rectangular region. If N blocks can be placed adjacently in the region, then the resistance in terms of the sheet resistance, $R_\square$, is given by

$$R = NR_\square \tag{2.2-4}$$

An example illustrating this technique is depicted in Fig. 2.2-1*c*.

Occasionally it is necessary to determine the resistance of nonrectangular regions such as that shown in Fig. 2.2-1*d* or the serpentined pattern of Problem 2.11. The problem of determining resistance of irregular regions is difficult, but for rectangular regions containing the right angles shown, the rule of thumb of adding 0.55 squares for each corner is often used.

Example 2.2-2. If $R_\square = 45\Omega/\square$, determine the value of the U-shaped resistor of Fig. 2.2-1*d*.

Solution. The U-shaped resistor is broken into squares of length W in the same figure. Counting squares and assigning 0.55 squares to each corner, we obtain a resistor of value 8.1 $R_\square = 364.5\ \Omega$.

It is often necessary to specify the temperature characteristics of resistors and capacitors. The Temperature Coefficient of Resistance (TCR) and Temperature Coefficient of Capacitance (TCC) are typically used for this purpose. These temperature coefficients, which are generally expressed in terms of ppm/°C, are defined by

$$\text{TC} = \left(\frac{1}{x}\right)\left(\frac{dx}{dT}\right) 10^6 \text{ppm/°C} \tag{2.2-5}$$

where x is the temperature-dependent value of either the resistor or the capacitor.

If the temperature coefficient is independent of temperature, then the value of the component at a temperature T_2 can be obtained from its value at temperature T_1 by the expression

$$x(T_2) = x(T_1)e^{[\text{TC}(T_2 - T_1)/10^6]} \tag{2.2-6}$$

which is often closely approximated by

$$x(T_2) \simeq x(T_1)[1 + (T_2 - T_1)(\text{TC}/10^6)] \tag{2.2-7}$$

If the TC is a function of temperature, then the previous expressions are good only in local neighborhoods of T_1. The value of TC, which can be either positive or negative, is determined by the material properties and is often quite small. The absolute values of TC are often less than 1000 ppm/°C. Unfortunately, since integrated circuits are often expected to operate over a relatively wide temperature range (0–70°C commercial or −55 to 125°C military) the effects of the TC can be significant.

Example 2.2-3. Determine the percentage error introduced in calculating $x(T_2)$ from $x(T_1)$ by using the simpler linear equation (2.2-7) compared to (2.2-6) if $T_1 = 30°\text{C}$, $T_2 = 60°\text{C}$, and the temperature coefficient is constant and given by TC = 1000 ppm/°C.

Solution. Using the exact expression we obtain

$$x(60) = x(30)e^{(.03)} = x(30)(1.0304545)$$

Using the approximate expression we obtain

$$x(60) = x(30)[1 + (30)(.001)] = x(30)(1.030000)$$

Thus, the error is about 0.044%.

Some resistors and capacitors, in addition to being temperature dependent, are also somewhat voltage dependent. This voltage dependence introduces nonlinearities in circuits using these devices along with the corresponding harmonic distortion (THD) in many applications. The voltage dependence of resistors and capacitors is characterized by the voltage coefficient, defined by

$$\text{VC} = \left(\frac{1}{x}\right)\left(\frac{dx}{dV}\right)10^6 \text{ppm/V} \tag{2.2-8}$$

where x is the voltage-dependent value of a resistor or capacitor. The voltage coefficient is analogous to the temperature coefficient, as can be seen by comparing Eqs. 2.2-5 and 2.2-8.

2.2.1 MOS Processes

A brief qualitative discussion of the principle of operation of the MOS transistor at dc and low frequency is now presented to provide insight into the MOS process itself. A detailed quantitative presentation about modeling these devices appears in Chapter 3.

OPERATION OF THE MOSFET. Consider the n-channel enhancement MOSFET shown in Fig. 2.2-2. In the cross-sectional views it can be seen that the gate (Metal

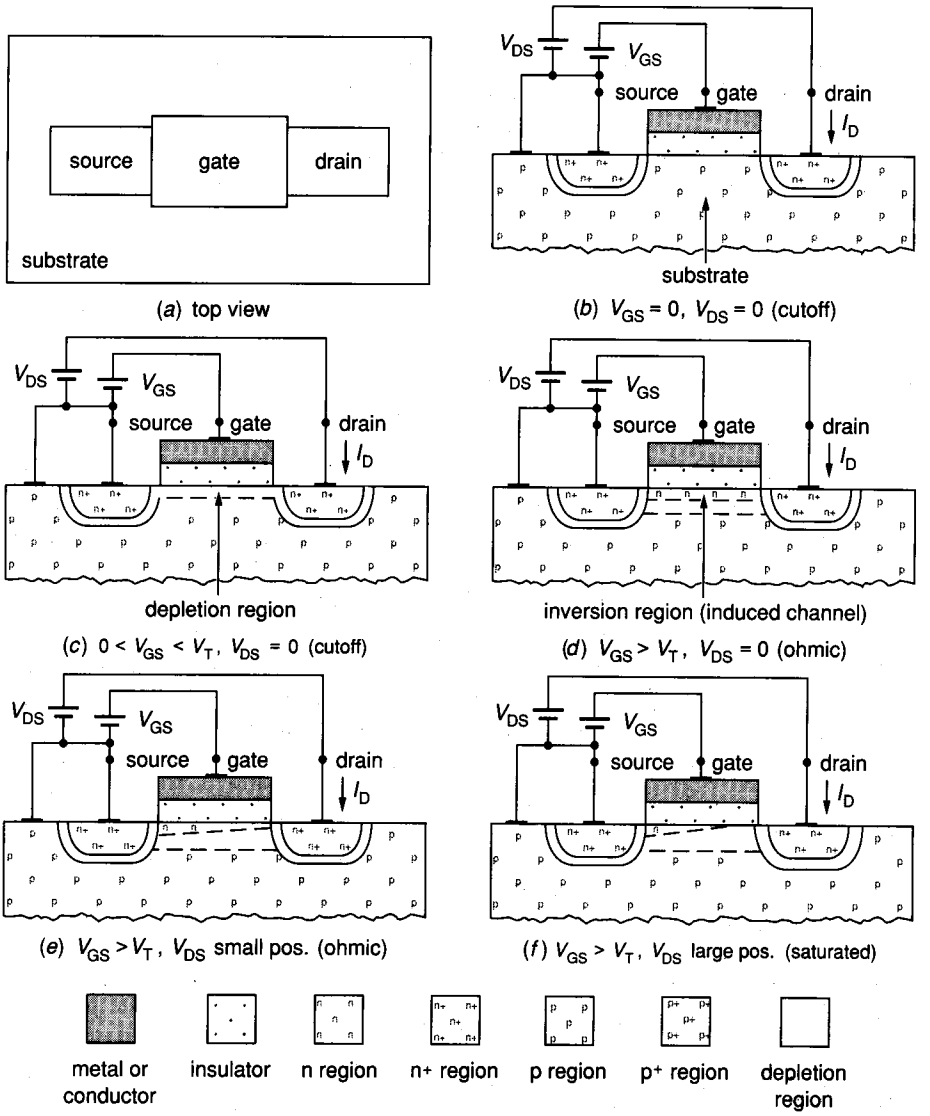

FIGURE 2.2-2
Operation of n-channel MOSFET (horizontal and vertical scale factors are different).

or conductor) is over the insulator (Oxide), which is in turn over the substrate (Semiconductor). The source of the acronym MOSFET should be apparent. If the substrate is tied to the source as shown in Fig. 2.2-2*b*, then with a zero gate–source voltage the n-type drain and source regions are isolated from each other by the p-type substrate, preventing any current flow from drain to source. A depletion region also forms between the n^+ drain and source regions and the lightly doped substrate, as depicted in Fig. 2.2-2*b-f*. The corresponding pn junction is reverse

biased under normal operation and has minimal effects on current flow at dc and low frequencies.

If a positive gate voltage is applied, electrons will start to deplete the substrate near the surface under the gate. This tends to deplete the p-type substrate in this region and form what is called a depletion region under the gate. A simplified pictorial presentation of this situation is shown in Fig. 2.2-2*c*.

If the gate voltage is increased sufficiently, a number of electrons will be attracted to the substrate surface under the gate sufficient to make this region n-type. This n-type region, which is created electrically (by the electric field established by the gate bias) in the p-type substrate, is called an *inversion layer*. The gate–source voltage necessary to create the inversion layer is called the *threshold voltage*, V_T. The inversion layer, shown in Fig. 2.2-2*d*, is often termed the *channel* of the MOSFET.

Once the inversion layer is created, current will flow from the drain to source or source to drain if a small voltage is applied between these regions. The insulator under the gate prevents any gate current from flowing, thus forcing the current entering the drain to be equal to that leaving the source. Increasing the gate–source voltage beyond the threshold voltage brings additional electrons under the gate, causing an increase in the thickness of the inversion layer and increased current flow from the drain to source (or source to drain) under a fixed drain-to-source bias. If a large drain–source (or source–drain) voltage is applied, this voltage itself will tend to deplete the inversion layer due to the potential drop across the region caused by the current flow. A cross section of the device is shown under a small drain-to-source bias in Fig. 2.2-2*e* and under a large drain-to-source bias in Fig. 2.2-2*f*. For a fixed gate–source voltage, there is a value of V_{DS} that effectively pinches off the channel near the drain. This does not cause a decrease in drain current (I_D), for if it did, the inversion layer would immediately reappear since the channel current itself causes the pinching of the inversion layer. If the value of V_{DS} is increased further, the drain current will remain nearly constant.

When the gate–source voltage is greater than the threshold voltage, the MOS transistor is said to be operating in the *ohmic region* prior to the pinching of the channel and in the *saturation region* when the channel is pinched off. If the gate–source voltage is less than the threshold voltage, almost no drain or source current will flow even when a bias is applied to the drain and source contacts. In this case the device is said to be *cutoff*. The relationship between I_D, V_{DS}, and V_{GS} for a typical MOSFET, termed the output characteristics, is shown in Fig. 2.2-3, along with the ohmic and saturation regions of operation. The cutoff region is the $I_D = 0$ line in this figure. The gate current remains at $I_G = 0$ in all three regions of operation.

The value of the threshold voltage is determined by the concentration of the p-type impurities in the substrate. If some n-type impurities are added to the region under the gate near the surface of the substrate, the threshold voltage will decrease. If sufficient impurities are added, the region itself will become n-type and the threshold voltage will become negative. An n-channel device with

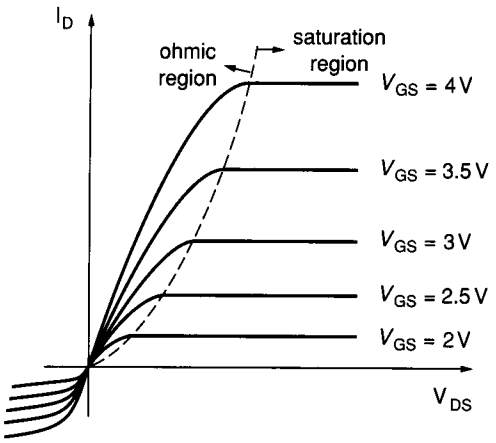

FIGURE 2.2-3
Typical output characteristics for an n-channel MOSFET.

a positive threshold voltage is termed an *enhancement MOSFET* and those with a negative threshold voltage are termed *depletion MOSFETs*. MOS devices formed in a p-substrate (or tub) and thus having n-type drain and source diffusions and an n-type channel are termed *n-channel transistors*. Those formed in an n-type substrate (or tub) with p-type drain and source diffusions and a p-type channel are termed *p-channel transistors*. In contrast to the convention introduced above for n-channel transistors, p-channel transistors with a negative threshold voltage are termed enhancement devices and those with a positive threshold voltage are termed depletion devices.

Commonly used symbols for enhancement and depletion n- and p-channel devices are shown in Fig. 2.2-4. Since the polarity of the substrate is often known for either the NMOS or PMOS processes, the simplified notation shown in Fig. 2.2-4, which does not maintain this information in the device symbol, has been widely adopted for both NMOS and PMOS devices.

2.2.1*a* NMOS Process

Although both NMOS and PMOS processes are currently available, the NMOS process has been used more extensively in recent years. The NMOS process is preferred because the characteristics of the n-channel MOSFET are preferable to those of the p-channel MOSFET. This is attributable to a higher mobility for electrons than for holes. The discussion that follows will be based upon the NMOS process. Modifications of this presentation to describe the PMOS process are straightforward and are thus left to the reader.

A discussion of a generic double-polysilicon, self-aligned silicon gate NMOS enhancement/depletion process follows. This can be considered as a typical standard process although processes that offer more as well as less flexibility are also standard. This process is similar to a widely used MOSIS NMOS process augmented by a second polysilicon layer. The same basic approach used

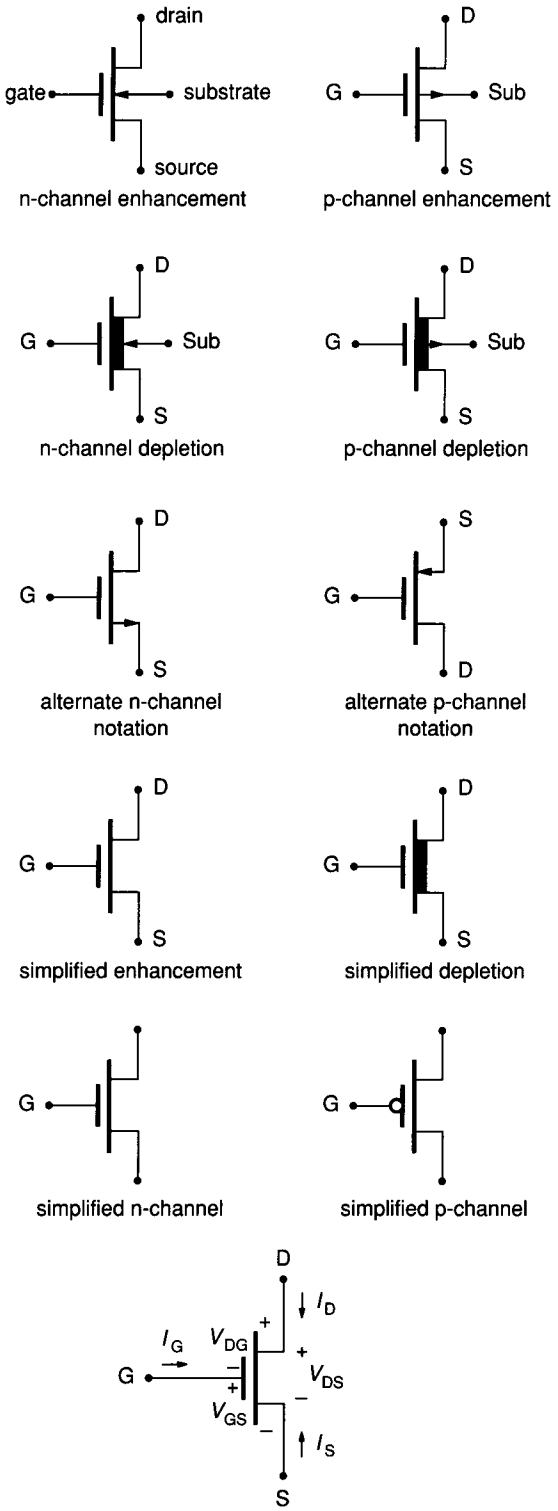

FIGURE 2.2-4
Symbols for MOS transistors.

here is used for other NMOS processes. The MOS transistors in this process are similar to those depicted in Fig. 2.2-2, with the exception that polysilicon (a good conductor) is used instead of metal for the gate. Field effect transistors with polysilicon gates are also called MOS transistors or MOSFETs even though the acronym MOS is no longer completely descriptive.

The devices that are available in this process are

1. n–channel enhancement MOSFETs.
2. n–channel depletion MOSFETs.
3. Capacitors.
4. Resistors.

The method of physically constructing each of these components in this process and interconnecting them to form the simple circuit shown in Fig. 2.2-5 will now be addressed. Both top views and cross-sectional views are presented in Fig. 2A.1 of Appendix 2A. Cross sections are along sections AA′ and BB′ of Fig. 2A.1*a*. A summary of the major process steps appears in Table 2A.1 of Appendix 2A. Additional information about this process relating to layout sizing rules, physical feature sizes, and electrical characterization parameters of the generic NMOS process can be found in Tables 2A.2–2A.5 of Appendix 2A. (Table 2A.3 appears in Plate 5 in the color plate insert.) The circuit designer must present the top view of each mask level for fabrication but must have a firm understanding of the cross-sectional view for effective design.

For the NMOS process, the starting point is a polished p-type silicon disc. The thickness of the disc is typically around 500 μ. A layer of SiO_2 in the neighborhood of 1000 Å thick is first added to the entire wafer using the oxidation process. On top of this a layer of Si_3N_4 (about 1500 Å thick) is applied by the

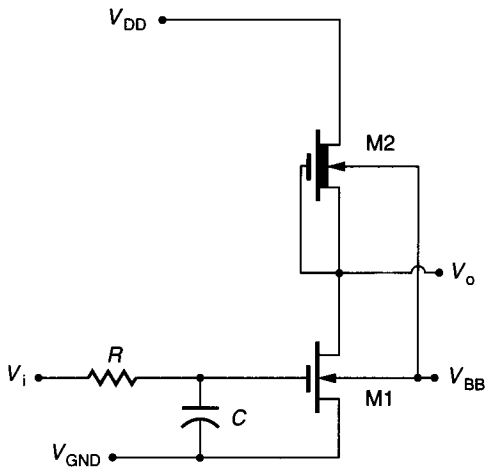

FIGURE 2.2-5
A simple NMOS circuit.