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

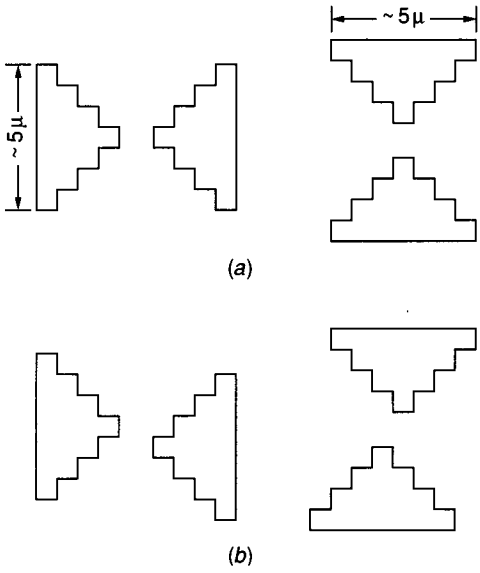


FIGURE 2.1-1
Alignment marks: (*a*) Mask marks, (*b*) Simulated positioning of alignment marks after fabrication.

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.

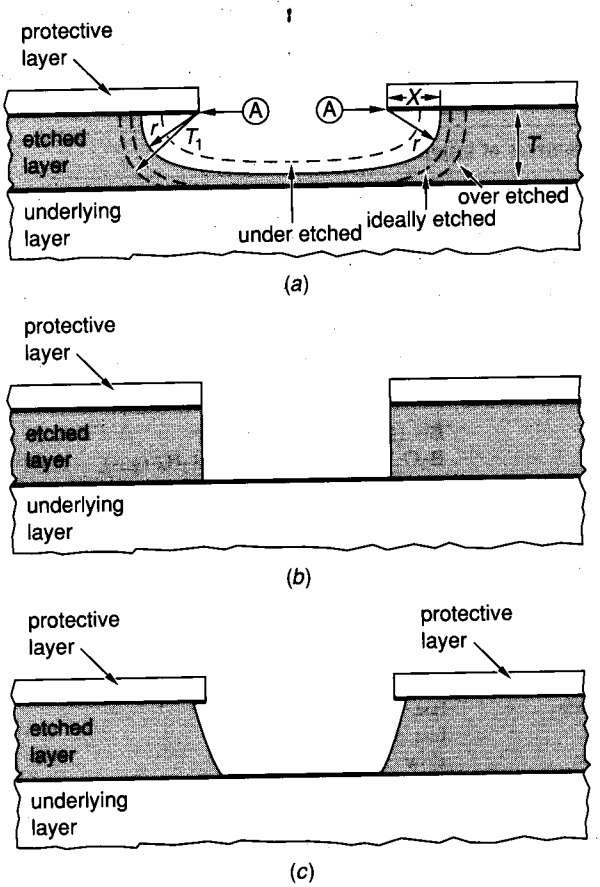


FIGURE 2.1-2
Characteristics of etches: (*a*) Isotropic etch, (*b*) Anisotropic etch, (*c*) Preferential etch.

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.

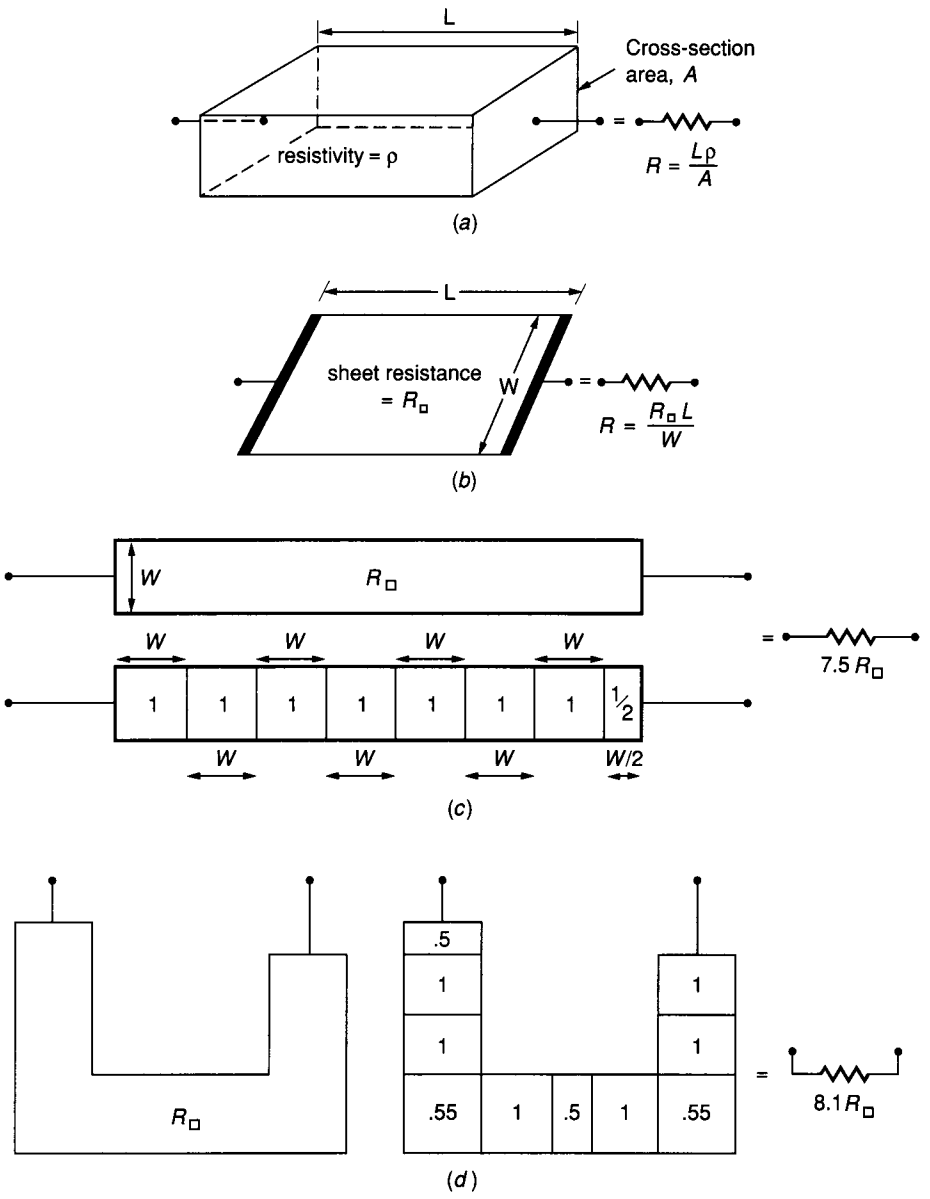


FIGURE 2.2-1
Resistive characteristics of bulk and sheet materials: (*a*) Resistivity, (*b*) Sheet resistance, (*c*)&(*d*) Graphical calculations from sheet resistance.

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

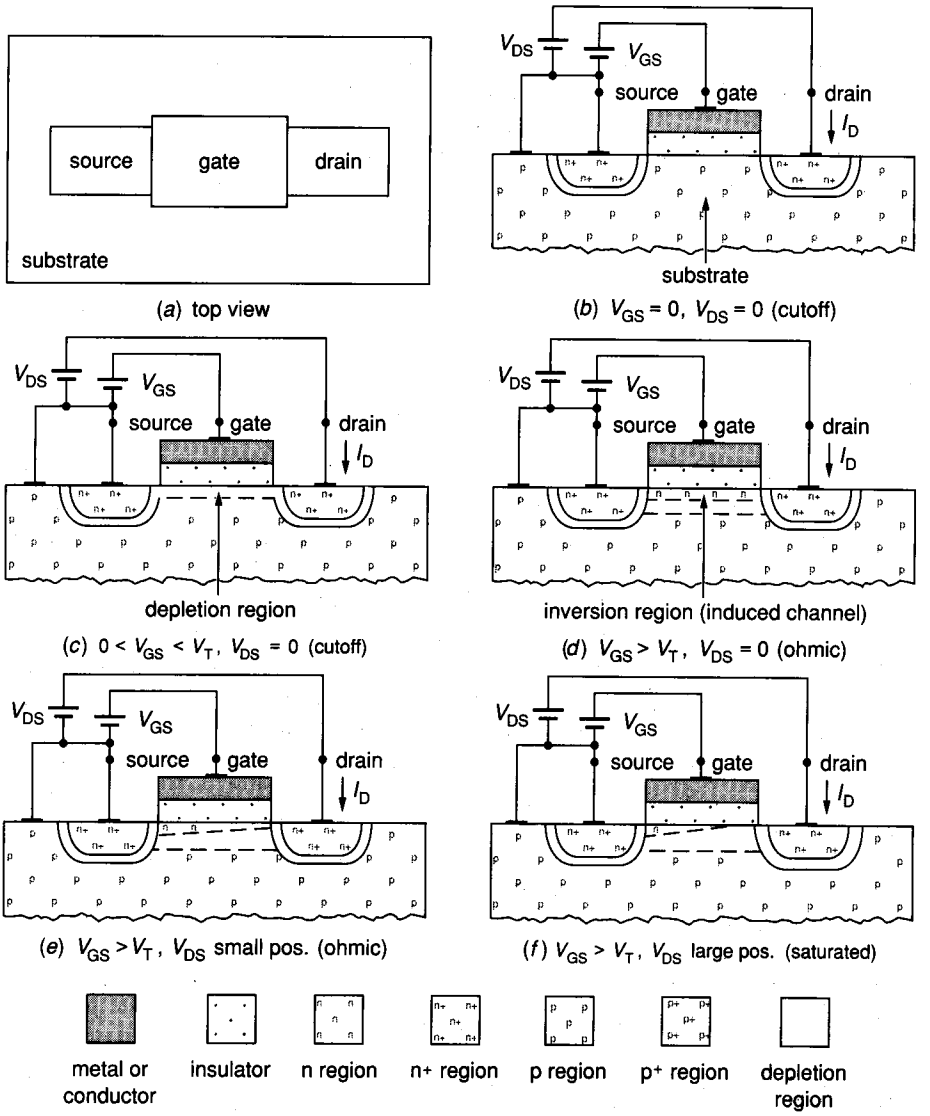


FIGURE 2.2-2
Operation of n-channel MOSFET (horizontal and vertical scale factors are different).

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

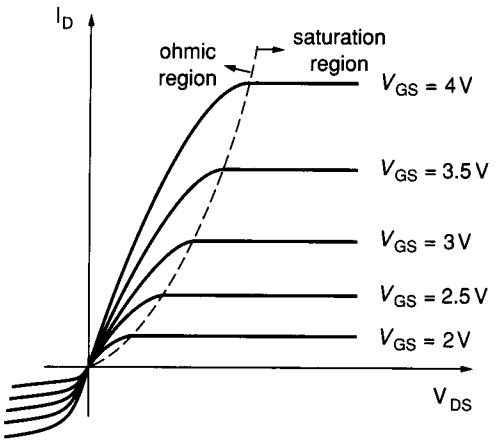


FIGURE 2.2-3
Typical output characteristics for an n-channel MOSFET.

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

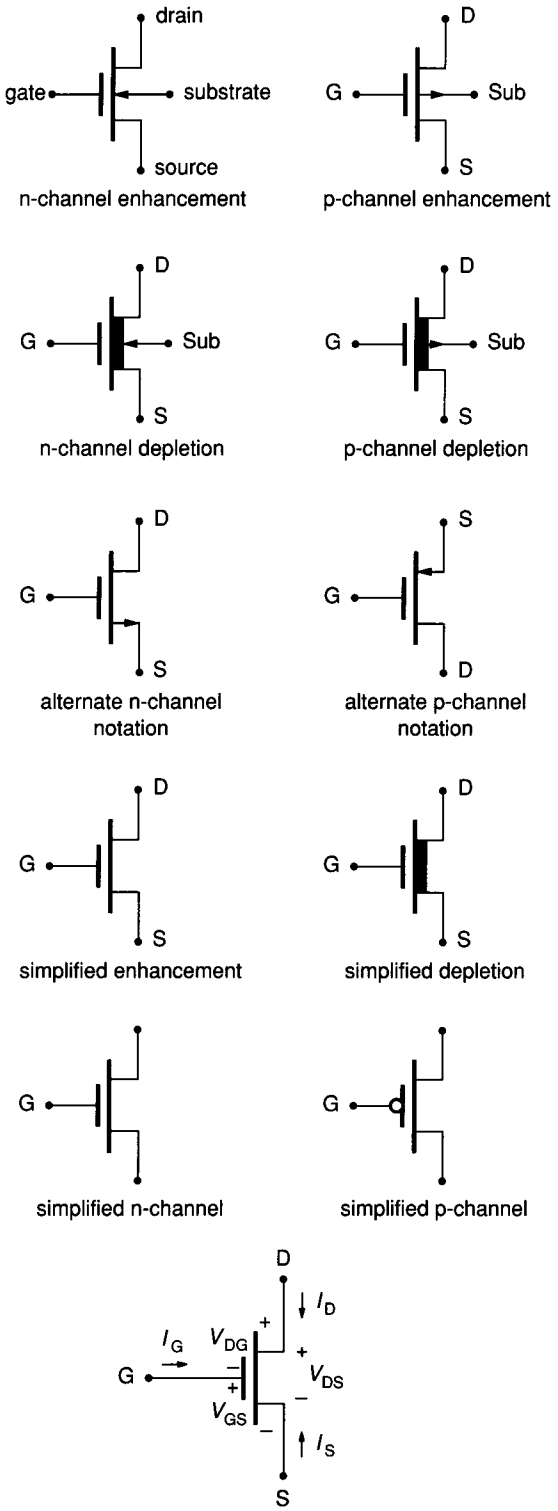


FIGURE 2.2-4
Symbols for MOS transistors.

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

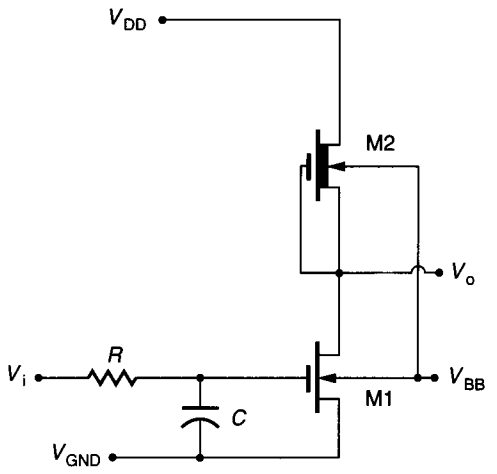


FIGURE 2.2-5
A simple NMOS circuit.

CVD process. Following the application of a layer of photoresist, Mask #1 is used to pattern the surface. Mask #1, which is often called the *moat*, or n^+ *diffusion* mask, defines in photoresist the drain, source, and channel regions of all transistors as well as any other regions where n^+ implants are desired. After exposure, development removes the photoresist layer in areas that are not to be moat (i.e., the *complement of the moat*, or the *antimoat*). A top (Mask #1 pattern) and cross-sectional view at this stage of what will be the two transistors appear in Fig. 2A.1*a*. The Si_3N_4 is then etched from the areas not protected by the photoresist. A high-energy implant of p-type impurities (typically boron) is then applied to the entire wafer. The remaining photoresist protects the moat regions from this implant. This heavy implant is used to raise the threshold voltage in the antimoat region (often called the *field*) and to provide electrical isolation between adjacent devices. After this field implant and a drive in diffusion, the remaining photoresist is stripped. A thick layer of SiO_2 (about 10,000 Å) is then thermally grown by the oxidation process over the wafer. This layer is formed in the field, but no oxidation can take place in the region protected by the Si_3N_4 because Si_3N_4 does not oxidize. The thick field oxide layer is termed a local oxidation layer and is often called *LOCOS*. The oxidation consumes some of the substrate silicon. The second cross section in Fig. 2A.1*a* shows the state of the wafer following growth of the field oxide. This corresponds to Step 10 in the process scenario of Table 2A.1. Following removal of the Si_3N_4, the thin layer of SiO_2 under the Si_3N_4 is stripped and another SiO_2 layer is grown.

With the moat now protected only by the very thin SiO_2 layer, a light n-type implant over the entire wafer can be applied (optional) to set the threshold voltage of the enhancement devices. This implant is light enough so that all p-type regions remain p-type. If used, the implant is applied to the entire wafer to avoid the need for an additional mask.

A heavier selective implant is required in regions that are to serve as the channels of depletion transistors. To achieve this, a second layer of photoresist is applied to the entire wafer, and the second mask, Mask #2 (termed the *implant* mask), is used to pattern the photoresist so that only the channel regions of depletion transistors are unprotected. Another n-type implant is used to make the exposed regions n-type with the remaining photoresist serving as an implant mask. After stripping the photoresist, the wafer is as shown in Fig. 2A.1*b*. This corresponds to the status of the wafer at Step 19 in Table 2A.1.

Direct contacts between the lower polysilicon layer and moat are termed buried contacts. In the process scenario of Table 2A.1 this is listed as an optional step and is not used in the layout shown in Fig. 2A.1 although it could be used, if available, to reduce the area required for contacting the gate of M2 to its source. To make a buried contact, the thin gate oxide must be patterned and etched to remove the insulating SiO_2 layer and create paths (vias) through which the following polysilicon layer can contact the moat. Mask #A is used to pattern the buried contact vias. Although the buried layer contact can reduce area, the additional processing costs needed to provide this feature are often not justified; hence the buried contact feature is often not available in NMOS processes.

After stripping of any thin oxides that may be present at this stage in the moat region, a uniform thin layer of SiO_2, often termed *gate oxide* (200 to 1000 Å thick), is grown on the surface of the wafer. Stripping and regrowing provide better control of the critical gate oxide thickness. A layer of polysilicon (termed POLY I), which is about 2000 Å thick, is then deposited on the surface of the entire wafer. This is covered with photoresist, patterned with Mask #3, and etched to remove unwanted POLY I. The POLY I layer is used as gates for both enhancement and depletion transistors, as a plate for capacitors, as a conductor, and for resistors. The formation of a capacitor, a resistor, and enhancement and depletion transistors can be seen in Fig. 2A.1*c*. This corresponds to Step 27 in the process scenario of Table 2A.1. Note that the POLY I layer is over gate oxide in cross section AA′ and above field oxide in cross section BB′.

The remaining uncovered thin layers of SiO_2 are stripped and another thin layer (500–1000 Å) of SiO_2 is again grown over the entire surface. This serves as the dielectric for POLY I–POLY II capacitors, as an insulator for POLY I–POLY II crossovers, and as the gate oxide for transistors that use the POLY II layer as the gate. The thickness of this oxide layer is often ideally the same as that of the first gate oxide layer. By stripping the unexposed oxide and regrowing, a buildup in the depth of the oxide layers is prevented and more uniformity is attained. A second layer of polysilicon (termed POLY II) is then deposited, followed by photoresist and patterning with Mask #B. In the circuit shown in Fig. 2.2-5, the POLY II layer is used only for the upper plate of the capacitor, as shown in Fig. 2A.1*d* (Step B.9 of Table 2A.1). Note that it is slightly smaller than the underlying POLY I layer. This difference is standard practice when trying to accurately match capacitors to make one plate a little smaller than the other so that the smaller plate will effectively define the capacitor area independent of slight misalignments of the two plates. Additional practical considerations that further improve matching will be discussed later.

After stripping the thin SiO_2 layers, an n^+ diffusion is applied to the entire wafer. The field oxide and polysilicon layers serve as masks to the diffusion and prevent impurities from reaching the substrate in the protected areas. The n^+ diffusion creates the n-type drain and source regions of all transistors and makes any other unprotected moat areas n-type. The portion of the light depletion diffusion that is not protected by the polysilicon gates also becomes more heavily doped. Note that although no mask is used at this step, the n^+ diffusion mask, which was used as the first masking step, has essentially determined the n^+ diffusion regions fabricated at this step. The n^+ diffusion also penetrates into any exposed polysilicon layers, increasing the conductivity in these regions. The n^+ diffusion depth is about 5000 Å. It will be seen in the section discussing process parameters that the sheet resistance for POLY I and POLY II layers is about the same but that the sheet resistance of POLY I under POLY II is higher. This is due to the absence of the additional n^+ diffusion in the lower layer. Since the polysilicon gates serve as masks for the n^+ drain and source diffusions, the process is said to be *self-aligned*. In a self-aligned process, small misalignments of the gate (POLY I or POLY II) masks will not affect the gate geometry or dimensions, nor will they make the transistors nonfunctional.

Next, another insulating layer is deposited over the wafer surface. Doped deposited oxide, such as PSG, is often used for this purpose. This rather thick layer, ~ 6000 Å, serves as an insulator between the uppermost polysilicon layer and the subsequent metal layer. The field oxide depth is further increased with this deposited oxide layer. The entire wafer is again covered with photoresist, and Mask #4 (actually the fifth or sixth mask if POLY II and/or buried contact options are available) is used to pattern *contact openings* for the purpose of obtaining electrical contact from the top with the desired components. After the photoresist is developed, an etch that attacks the insulating layer but does not affect polysilicon or silicon makes the required openings. This etch is stopped in the vertical direction only by polysilicon or the single-crystal silicon of the substrate. The wafer takes the form shown in Fig. 2A.1*e* (Step 34 of Table 2A.1).

Metal (typically aluminum) is then deposited over the entire wafer, followed by another layer of photoresist. This metal layer is typically about 7000 Å thick. This photoresist is patterned by Mask #5, followed by an etch to remove unwanted metal. The metalization is used to interconnect components and provide external access to the integrated circuit. A metalization that interconnects the four basic components to form the circuit shown in Fig. 2.2-5 is shown in Fig. 2A.1*f* (Step 40 of Table 2A.1).

Large, square metal areas, called *bonding pads*, are needed to allow for contact with the IC package. Small bonding wires will later be connected from these pads to the pins on the IC package. These pads are also patterned with the metalization mask but are not shown in Fig. 2A.1*f* because of the large amount of area required for bonding pads relative to that needed for the components in Fig. 2.2-5.

A bonding pad is shown in Fig. 2A.1*g*. Four of these would be needed to interface the circuit of Fig. 2.2-5 with the IC package. The V_{BB} contact comes from the bottom side of the substrate. The bonding pad size has remained relatively constant for a long period of time even though considerable reductions in feature size of geometries on the die itself have been experienced. This is because the methods of physically mounting the die in packages and interconnecting the bonding pads to the pins in the package have not changed much. With bonding wires typically about 1 mil in diameter, it is difficult to reduce bonding pad size significantly.

The entire surface is finally covered with a passivation layer (often called glass or p-glass) to provide long-term stability of the IC by minimizing atmospheric contamination. A layer about 10,000 Å thick is often used for passivation. Since this layer is also an electrical insulator, it is necessary to again pattern it and make openings above the metal pads to allow for attaching the bonding wires. The final mask, Mask #6, is used for this purpose and is shown in Fig. 2A.1*g*.

For the simple circuit of Fig. 2.2-5, the area required for the bonding pads dominates that needed for the circuit itself. For simple circuits this is generally the case but as the complexity of the circuit increases, the percentage of the total area required for bonding pads becomes quite small.

Giving the information for each mask separately, as was done in Fig. 2A.1, makes it difficult to perceive the entire circuit and determine layer to layer

alignment. Sophisticated software packages termed layout editors are widely used, in which layers are color-coded and displayed simultaneously on high-resolution monitors. A single layout that simultaneously shows all mask information is shown in color in Plate 2. A color convention has been established for distinguishing separate layers. The color convention adopted in Plate 2 corresponds to that used in the MOSIS process and is discussed in more detail later in this chapter.

An interesting observation can be made from the layout of the MOS transistors of Fig. 2A.1. The MOS devices are totally geometrically symmetric with respect to drain and source and so must also be electrically symmetric. The designation of *drain* and *source* is thus arbitrary. In many applications a convention has evolved for convenience and consistency in device modeling in regard to drain and source designation. This convention will be discussed in Chapter 3. When appropriate, we will follow the established convention throughout this text.

In the process described, several alternative methods for constructing resistors and capacitors are available. For example, a region of moat with two contacts can be used as a resistor, and a capacitor can be made between POLY I and metal.

Processing step modifications such as omission of one polysilicon layer, omission of the depletion mask, substitution of metal gates for the polysilicon gates, and addition of another mask and implant to create enhancement transistors with two different threshold voltages are possible and common. Process procedure modifications such as using diffusions instead of implants; changing types of impurities; varying the thickness of oxide, polysilicon, or metal layers; including or excluding oxide stripping and regrowing steps; and changing types of photoresist are widespread and play major roles in yield, fabrication costs, and performance. The IC design engineer must be familiar with the process steps that will be used in fabrication when embarking on a new design.

For the process just described, it is the responsibility of the circuit designer to provide all information necessary to construct the seven masks shown in Fig. 2A.1. The size, shape, and spacing of the components are judiciously determined. The size and shape affect the performance of the circuit and are at the control of the circuit designer for optimizing performance, within constraints of minimum allowable size as determined by the capabilities of the process and maximum size as determined by economics. The spacing is also constrained by the capabilities of the process itself. The spacing and sizing specifications are obtainable from the design rules of the process, which are discussed later in this chapter. A process engineer will typically be responsible for providing design rules for a particular process.

2.2.1*b* CMOS Process

A discussion of a typical generic single-polysilicon silicon gate, p-well, n-substrate CMOS process follows. As in the NMOS case, variants in this process—such as a second metal layer, a second polysilicon layer, additional implants, oppositely doped substrate, or metal gates—are also well established. The devices available in the CMOS process under consideration are

1. n-channel MOSFETs.
2. p-channel MOSFETs.
3. Capacitors.
4. Resistors.
5. Diodes.
6. npn bipolar transistors.
7. pnp bipolar transistors.

The diodes and bipolar transistors are often considered parasitic components and are generally not extensively used as components in the circuit design itself. The process is tailored to maintain optimal characteristics in the n- and p-channel MOSFETs at the expense of poor characteristics for the bipolar transistors.

A method of physically constructing each of the first four components in the list will be considered. The approach followed here is similar to that followed for the NMOS process except that all mask details are included on the single layout of Fig. 2B.1 of Appendix 2B. A color version of this figure appears in Plate 3. Fewer details about oxide growth, photoresist application and patterning, and so on are provided since these steps are very similar to the corresponding steps for the NMOS process. Cross-sectional views along AA′ and BB′ in Fig. 2B.1*a* after each major step are shown in Fig. 2B.1. Interconnections follow the approach used in Section 2.2-1*a* for the NMOS process and are not discussed here. A summary of the major process steps appears in Table 2B.1. Additional information about this process relating to layout sizing rules, physical feature sizes, and electrical characterization parameters of the generic CMOS process can be found in Tables 2B.2–2B.5 of Appendix 2B. The process described here is very similar to the 3 μ CMOS/bulk process available through MOSIS.[7]

The starting point of this CMOS process is a polished n-type silicon disc. A layer of SiO_2 is first grown on the entire disc, followed by the application of a layer of photoresist. This photoresist is patterned with Mask #1 to provide openings for a p-tub (alternatively, p–well), which will serve as the substrate for the n-channel MOS devices. Either a deposition or implant is used to introduce the p-type impurities that form the tub. This diffusion is quite deep (about 30,000 Å). The remaining photoresist and SiO_2 are then stripped, a thin layer of SiO_2 regrown, and the entire surface covered with a layer of Si_3N_4.

Mask #2, termed the *moat mask* or the *active mask* by MOSIS, is used to pattern the Si_3N_4 layer. The Si_3N_4 layer is then etched away except above the regions that are to be the n^+ and p^+ diffusions or channel regions for the n-channel and p-channel MOSFETs. These diffusions will be added by subsequent processing steps to form drain and source regions for MOSFETs as well as to form guard rings. These protected regions are again termed moat. After the Si_3N_4 layer is opened, the remaining photoresist is stripped. Figure 2B.1*b* depicts the wafer after Step 15 of the process scenario of Table 2B.1.

An optional field threshold adjust step may be introduced at this point. This field threshold adjust would be used to raise the threshold voltage in the n-type

substrate in regions that will not contain devices. This will provide increased isolation between the p-channel transistors. Although an additional mask is required for this field adjust (Mask #A1 of Table 2B.1), the mask would be the complement of the union of the p-well mask and the active mask, Masks #1 and #2. As such, this mask information would be generated automatically and need not be separately provided by the designer.

A thick layer of field oxide (typically 10,000 Å) is grown in the regions not protected by the remaining Si_3N_4 that was patterned with Mask #2. The Si_3N_4, along with the remaining SiO_2 that was under this layer, are then stripped. The wafer at this stage is as depicted in Fig. 2B.1*c*. This corresponds to Step 18 in the process scenario of Table 2B.1. Note that along cross section AA′ several isolated areas are not protected by the field oxide. These areas will be used for fabricating transistors and guard rings. No breaks in the field oxide appear in the BB′ cross section. This corresponds to the region where the resistor and capacitor will appear; these devices are fabricated on top of the field oxide.

Next, a thin, uniform layer of SiO_2 (200 to 1000 Å), called in this case gate oxide, is regrown. A layer of polysilicon (typically 2000 Å) is then deposited, covered with photoresist, and patterned with Mask #3. This polysilicon layer, termed POLY or POLY I, is used for the gates of all transistors, as a plate on capacitors, for resistors, and for interconnects. Following etching and stripping, the wafer takes the form shown in Fig. 2B.1*d*. This corresponds to Step 25 in the process scenario of Table 2B.1.

An optional second polysilicon layer, termed POLY II, could be included here, as provided in the process scenario. The second polysilicon layer is not depicted in Fig. 2B.1. This second polysilicon layer would be separated from the first by a thin (500 to 1000 Å) insulating layer of SiO_2. The main purpose of the second polysilicon layer would be for the formation of capacitors with POLY I and POLY II as electrodes, although this layer would also find some use in interconnects and crossovers if available. The capacitance density and electrical characteristics of the poly–poly capacitors are more attractive than those obtainable with other capacitors available in this process. An additional mask, termed the POLY II or *electrode mask*, is needed to pattern this polysilicon layer. The etch of both the POLY I and POLY II layers produces an abrupt, sharp edge, making reliable coverage of this edge with thin material difficult. Since the oxide between POLY I and POLY II is thin, crossing of a POLY I boundary with POLY II may result in either a break in the POLY II or a shorting of POLY I and POLY II. To circumvent these problems, the crossing of a POLY I boundary with POLY II is often not permitted.

At this stage the drain and source diffusions for both the n-channel and p-channel transistors are added. Although two different types of diffusions and hence two separate masks are required, the designer need specify only one of the two masks. In this process, only those areas not protected by field oxide are capable of accepting any diffusion impurities. This is termed the moat, or active, region. It is further provided in this process that any moat area that is not exposed to n-type impurities will be exposed to p-type impurities. Consequently, the designer selects those moat regions that are to become p-type with Mask #4,

which is termed the p^+ select mask. The n^+ select mask (Mask #5), which is used to pattern those regions of moat that are to become n-type, is automatically generated from the complement of the p^+ select mask intersected with the moat (active) mask. p^+ select is used in the substrate to form p-channel transistors and interconnects and is used in the p-well to provide ohmic contact to the p-well as well as for guard rings.

Correspondingly, n^+ select is used in the p-well to form n-channel transistors and interconnects, and in the substrate to make top ohmic contacts as well as additional guard rings. Further comments about guard rings and their role in latch-up protection appear in Section 2.4. As was the case in the NMOS process, the polysilicon layer or layers are patterned prior to the p^+ and n^+ diffusions. The polysilicon that lies in the moat serves as a diffusion mask for these diffusions and provides self-alignment of the gate with the drain and source regions.

The n^+ and p^+ diffusions are much shallower than the p-well diffusion and are typically in the 5000 Å and 7000 Å ranges, respectively. Following the p^+ and n^+ diffusions, which occur prior to Step 36 in the process scenario, the cross-sectional profile is as shown in Fig. 2B.1*e*. The n-channel and p-channel transistors, along with the ohmic contacts and guard rings, are clearly visible at this stage. A thick insulating layer, which is a deposited oxide (often PSG), is then placed over the entire wafer. This insulating layer is about 6000 Å thick and serves as an insulator between the uppermost polysilicon layer and the subsequent metal layer. This causes a further thickening of the field oxide and is depicted above the dashed interface of the field oxide layer shown in Fig. 2B.1*f*.

Mask #6 is used to pattern contact openings. Areas unprotected by photoresist after patterning are etched away. This etch will consume insulating layers but is stopped by either polysilicon or the silicon substrate. This provides for metal contact of either a polysilicon layer or a p^+ or n^+ diffusion depending on which is the uppermost layer present.

After the contacts are opened, metal is applied uniformly to the wafer and patterned with Mask #7. This is termed the *metal mask* (or, if subsequent metal layers are to be added, the *metal 1 mask*). This corresponds to Step 48 in the process scenario and the cross section of Fig. 2B.1*f*.

An optional second metal can be added at this stage. This requires two additional mask steps, one for making contact with underlying metal 1 and the other for patterning the second metal layer. The mask used to pattern the contact openings, or vias, between the two metal layers is termed the *via mask*. Polyimide is often used as the insulating layer between the two metals because it offers advantages in step coverage over other commonly used insulating layers.

Following application of a passivation layer, often termed p-glass, the passivation is opened above the bonding pads to provide for electrical contact from the top with Mask #8. This is often termed the *glass mask*. The pad layout is similar to that discussed for the NMOS process and depicted in Fig. 2A.1*g*.

This completes the CMOS processing steps for the generic CMOS process scenario of Table 2B.1. The resistor, capacitor, n-channel MOSFET, and p-channel MOSFET should be apparent from Figs. 2B.1*a* and 2B.1*f*.

2.2.1*c* Practical Process Considerations

The equipment needed for the CMOS process is basically the same as is needed for the NMOS process previously described. With this equipment the minimum feature size for the CMOS process is comparable to that for the NMOS process. It should be noted that eight masks and considerably more processing steps than are required for the basic six-mask NMOS process are needed for this CMOS process. In addition, it will be seen later that considerably more area is required for the same number of devices in a CMOS process than in an NMOS process with the same feature size. The increase in size is due largely to the required size of the large p-tubs and the n- and p-type guard rings. These increases in area are, however, often offset by less complicated designs and/or the superior performance that is attainable with the CMOS process.

Several physical and processing-dependent material characteristics cause the physical MOSFET to differ from the ideal. The processing-dependent material characteristics will be considered first.

WIDTH AND LENGTH REDUCTION. A typical cross section of the n-channel MOSFET along EE′ and FF′ of Fig. 2B.1*a* is compared with the ideal in Fig. 2.2-6. These cross sections are intentionally not to scale so that they will better illustrate the actual characteristics.

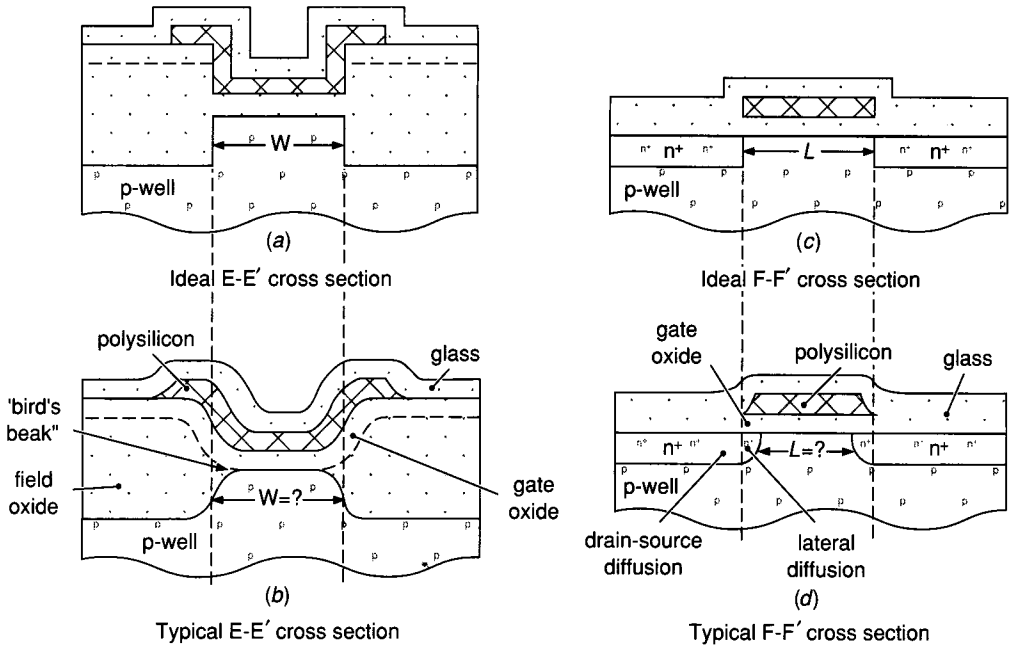


FIGURE 2.2-6
Width and length reduction in MOSFETS.

It will be seen later that the width and length of the MOSFET are key parameters at the control of the designer that play a major role in device performance. The width, W, is the width of the moat, or active, region as depicted in Fig. 2.2-6*a*, which corresponds to the EE′ cross section of the MOSFET, and the length L is the distance between the drain and source diffusions, as indicated in the FF′ cross section of Fig. 2.2-6*c*. It should be emphasized that the device dimensions are determined by the size of the *intersection* of the poly mask and the active mask and not by the dimensions of the poly pattern that forms the gate.

In the typical cross section of Fig. 2.2-6*b*, it can be seen that during the field oxide growth, encroachment into the active region effectively reduced the width of the transistor. This oxide encroachment is termed *bird's beaking* due to the distinctive shape of the encroachment. This is particularly troublesome because the width of the transistor is no longer precisely defined and because the exact amount of width reduction is not easily controllable. A second factor that affects the effective width is the accuracy with which the protective Si_3N_4 layer used to pattern the field oxide can be controlled. The effective width of this layer is affected by both the patterning of the photoresist and the problems associated with etching that were discussed in Section 2.1. In addition, since a thin SiO_2 layer (200–800 Å) is applied prior to the Si_3N_4 to minimize mechanical stress at the Si_3N_4 interface, the encroachment of the SiO_2 growth also limits accuracy.

Similar problems in controlling the length of the transistor exist, as indicated in Figs. 2.2-6*c* and *d*. The major source of length reduction is associated with the lateral diffusion of the drain and source diffusions, which are difficult to precisely control. Assuming the lateral and vertical diffusion rates are equal and that the diffusion depth is 5000 Å, the total length reduction due to lateral diffusion, since it diffuses in from both ends, would be around 1 μ. This is very significant and problematic in short channel transistors. Other factors that affect the effective length are the accuracy in patterning the photoresist that defines the polysilicon gate length and the accuracy in controlling the polysilicon gate etch itself.

In summary, both length and width reduction are inherent with existing processing technologies. Although they can be partially compensated for by considering these reductions during design or automatically adjusting (termed *size-adjust*) the geometrical database to over- or undersize the appropriate mask geometries, these effects are difficult to precisely control, and the exact width and length of the device are difficult to define. These effects are particularly troublesome for small geometries with device dimensions in the 1 μ or smaller range. Partial compensation with the mask size-adjust is often provided, thus allowing the designer to assume that the nominal value of the actual dimensions on silicon agree with those specified on the design.

LATERAL WELL DIFFUSION. Lateral diffusion associated with the creation of the p-well also deserves mention. The depth of the p-well is about 3 μ and, the lateral diffusion associated with the well formation is comparable. Although not a major factor limiting device performance, this lateral diffusion consumes considerable surface area and forces the designer to leave a large distance between isolated p-wells and between any p-well and p^+ diffusion in the substrate, thus

increasing chip cost by increasing die area. One way to partially minimize the impact of the large amount of area associated with well boundaries is to group large numbers of n-channel transistors into a single p-well when the wells for these transistors are to be tied to the same potential. Tradeoffs between these area savings and the corresponding increase in interconnect area must be made.

LATCH-UP. The physics of layered doped silicon is also problematic. It is well known that a four-layer sandwich of doped material, npnp or pnpn, forms a Silicon Controlled Rectifier (SCR). Once an SCR is "fired" (switched to on conducting state), it continues to conduct until the gate signal is removed and current flow is interrupted. Several parasitic bipolar transistors and an SCR are identified on the cross section of Fig. 2.2-7 which is based on Fig. 2B.1*f*. Several diodes can also be identified. Although this CMOS process has not been optimized for obtaining good performance of these bipolar devices, there are limited practical applications of some of the diodes and bipolar transistors. The SCR, however, is very undesirable and if it is caused to fire, excessive current will usually flow, causing destructive failure of the integrated circuit. The firing of the SCR is termed *latch-up* in CMOS circuits. The CMOS designer must make certain that latch-up cannot occur in any design.

Latch-up problems are strongly layout dependent. A theoretical treatment of the latch-up problem is beyond the scope of this text, but a thorough understanding of this problem is not needed for successfully designing CMOS circuits provided the designer is familiar with layout techniques that circumvent the problem. Guard rings are widely used to prevent latch-up. Exact requirements for guard ring

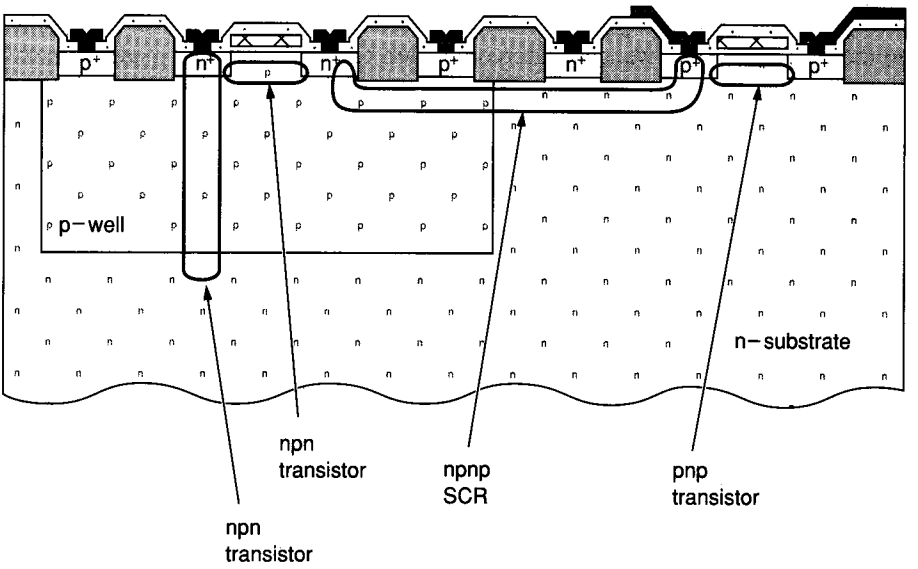


FIGURE 2.2-7
Parasitic transistors in a p-well CMOS process.

placement will be determined once a particular CMOS process is defined. In some processes, separate and additional n^+ and/or p^+ diffusions are included specifically for guard ring formation. This requires additional masks and additional processing steps. In the CMOS process discussed in this section, no additional masking or processing steps are required since the normal drain and source diffusions are also used to fabricate guard rings.

One way to obtain latch-up protection in the generic CMOS process of this section is to completely encircle every p-well with a p^+ guard ring. Such a guard ring is shown in Fig. 2B.1*a* around the periphery of the p-well. Metal contact is made as often as possible to this guard ring to further reduce resistance. The guard ring will then typically be connected via metallization to the lowest dc potential in the circuit—V_{SS} or ground, for example. Although not shown in Fig. 2B.1*a*, encircling the p-well with an n^+ guard ring provides additional protection and is also desirable. A partial n^+ guard ring separating the p-well from the p-channel substrate transistor can be seen in the same figure. As before, numerous metal contacts are made to this guard and it is subsequently tied to the highest potential in the circuit.

Breaks in a guard ring must be avoided since these breaks could provide a path for breakdown. Consider first the p^+ guard ring in the p-well. Breaks can occur one of two ways. The most obvious is to exclude a segment from either the moat mask or the p^+ select mask. The other way is to cross the guard ring *anywhere* with polysilicon in the process described in Table 2B.1. Such a crossing will cause a break because the polysilicon is patterned prior to the p^+ diffusion and serves as a mask to this diffusion. The break would thus occur under any polysilicon crossing of the intended guard ring. The exclusion of polysilicon crossing of the guard ring is undesirable from a circuit designer's viewpoint because it complicates interconnection between devices in the p-well and those outside the p-well; all interconnection crossings must be made of metal to avoid breaking the guard ring. Polysilicon crossing of guard rings in other process scenarios where a separate p^+ guard ring diffusion is available may be permissible. Correspondingly, breaks in the n^+ guard ring will occur if a segment is omitted from the active mask, if it is crossed with p^+ select, or if it is crossed with polysilicon.

Although complete enclosure of the p-well with the n^+ guard ring is desirable, some designers using the generic CMOS process described in this section use only the p^+ guard ring or have the p^+ guard ring and include the n^+ guard material only between the p-channel transistors and the p-well, as depicted in Fig. 2B.1*a*.

INPUT PROTECTION. Static breakdown is also of concern, and protection of inputs must be provided to prevent destructive breakdown when handling the devices. The major sources of concern are inputs that have a direct connection to a region separated from the rest of the circuit only by thin oxide, such as gate oxide or poly–poly oxide, with no direct connection to any diffused region. Such inputs would include the gates of any transistors or any connection to a floating polysilicon capacitor electrode. The breakdown is due to a destructive breakdown

of this thin oxide due to electric fields that exceed the oxide breakdown voltage. As stated in Chapter 1, silicon dioxide will break down when electric fields are in the 5 MV/cm to 10 MV/cm range. With 1000 Å gate oxides, this would occur for voltage inputs in the 50 V to 100 V range. Although these are beyond the maximum allowable input voltages specified for a typical 3 μ CMOS process, these voltages are much less than the static voltages experienced when handling these chips. The problem is even worse for thinner gate oxides. Such breakdown is destructive and must be prevented. Input protection circuitry is used for this purpose. This input protection must not interfere with the normal operation of the circuit. A single simple protection circuit is typically developed and is used repeatedly by connecting it to each pad that requires protection.

One common protection scheme involves connecting the input pads through a small polysilicon resistor to a reverse-biased diode that nondestructively breaks down at voltages below the critical gate oxide breakdown voltage. The node that is to be protected then becomes an internal node coincident with the node corresponding to the interconnection between the protection resistor and the diode. The resistor is used to safely limit peak current flow in the protection diode. In the CMOS process described in this section, this diode would be constructed by putting an n^+ diffusion in a p-well with a p^+ select guard ring around the periphery of the well. This guard ring would be connected to the lowest potential in the circuit, typically ground or V_{SS}, and the n^+ diffusion would be connected to the pad that is to be protected through the polysilicon resistor. This protection circuit provides protection through the reverse breakdown voltage of the diode if the input is positive and through normal forward-biased diode conduction if the input is negative. For the NMOS process, single diode protection can be attained by connecting the critical node through a polysilicon resistor to an n^+ diffusion. No guard ring is available or required in an NMOS circuit.

An alternative that provides all protection through normal forward-biased diode conduction is obtained if a second diode of opposite polarity shunts the diode just described. This diode is constructed from a p^+ diffusion in the substrate, with the n-substrate connected to the highest potential in the circuit and the p^+ diffusion connected to the intersection node of the first diode and the polysilicon resistor. It is recommended that this p^+ diffusion be encircled by an n^+ guard ring, which also would be connected to the substrate. Under normal operation the diodes in the input circuitry do not conduct, so the input protection is ideally transparent to the user. Actually, the diodes do contribute to a small amount of leakage current. They also contribute to a small parasitic capacitance connected to an ac ground, which may be of limited concern in some applications.

Although the input protection schemes discussed could be used on any input or output pad, such circuitry is generally not required on pads that are already directly connected to a diffusion region, even if they are also connected to layers separated by thin oxide from other nodes in the circuit, because the diffused region itself forms part of the diode and thus provides inherent self-protection. Nevertheless, care should always be exercised when handling any MOS devices, even if good circuit-level protection has been included, to reduce the chance of destroying the integrated circuit by static breakdown.

2.2.2 Bipolar Process

The basic active devices in the bipolar process are the npn and pnp transistors. These names are descriptive since the devices are constructed with three layers of n- or p-type semiconductor material, with the middle layer different from the other two. These layers can be fabricated either laterally or vertically. A simplified pictorial description of these transistors, including the established symbols for the devices, appears in Fig. 2.2-8. Several excellent references discuss the basic operation of the BJT.[3,12–17] The modeling of the BJT is discussed in Chapter 3. As will be seen later, the characteristics of the collector and emitter regions as well as their geometries are intentionally different and as such the designation of the collector and emitter contacts is not arbitrary. The convention that has been established for designating the collector and emitter contacts will be discussed in Chapter 3.

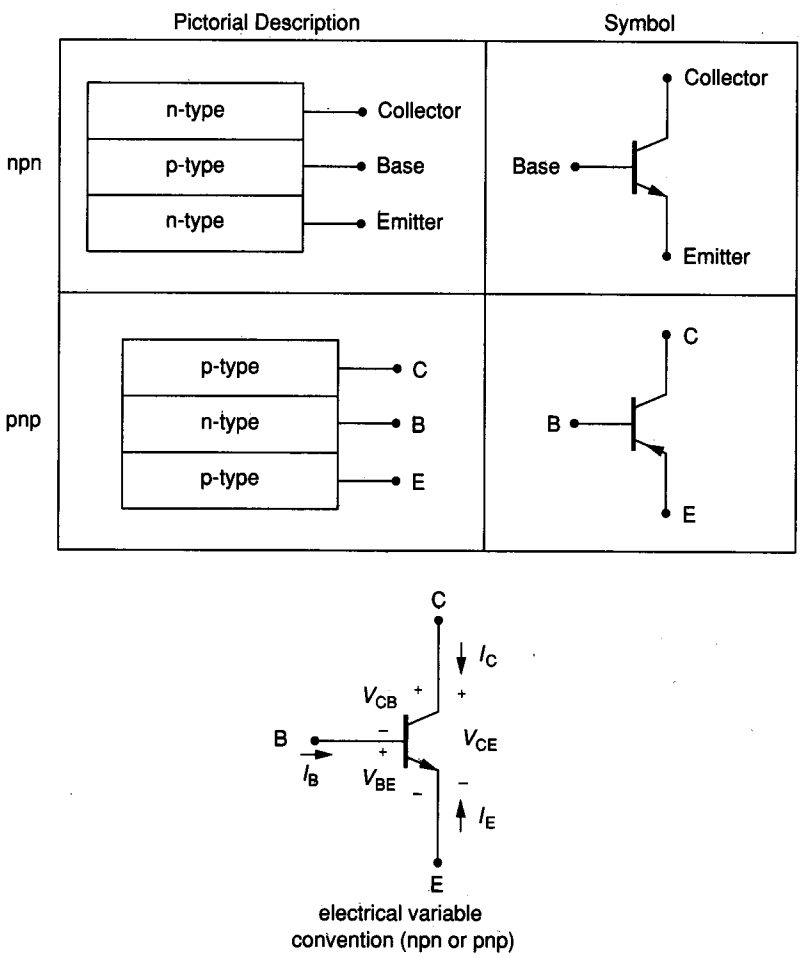


FIGURE 2.2-8
Bipolar transistors.

The components available in the bipolar process are

1. npn bipolar transistors.
2. pnp bipolar transistors.
3. Resistors.
4. Capacitors.
5. Diodes.
6. Zener diodes.
7. Junction Field Effect Transistors (JFETs)—not available in all bipolar processes.

A familiarity with the process is crucial to utilizing this wide variety of components in the design of an integrated circuit. Unlike discrete component circuit design with these same devices, whose characteristics can be specified over a wide range and which can be connected in any manner, the basic characteristics of the devices available for bipolar integrated circuits are determined by the process and the range of practical values and parameters is severely limited. In addition, the methods of interconnection are strictly limited, the basic devices have characteristics that are quite temperature dependent, and the passive component values are typically somewhat dependent on the signal applied. In spite of these restrictions (to be discussed later), very clever analog and digital bipolar integrated circuits have evolved. The bipolar process is used for the popular TTL, ECL, and I^2L digital logic families as well as a host of linear integrated circuits, including the 741 operational amplifier, the 723 voltage regulator, and the 565 phase locked loop. Although minor variances in the processing steps are common, the major differences are in device sizes and impurity concentrations and profiles. The discussion of a typical seven-mask bipolar process follows. The major process steps are outlined in Table 2C.1 of Appendix 2C.

The construction of npn transistors, pnp transistors, resistors, and capacitors will be considered. The location of these components can be seen in the top view containing mask information shown in Fig. 2C.1*a* of Appendix 2C.

The starting point in this bipolar process is a clean, polished p-type silicon wafer. A layer of SiO_2 is first grown over the wafer. Following application of a layer of photoresist, Mask #1 is used to pattern the n^+ buried layer. The n^+ buried layer serves the purpose of decreasing collector resistance and minimizing the parasitic current flow from collector to substrate in npn transistors. It also helps decrease the base resistance in lateral pnp transistors. Either a deposition or implant, followed by a drive in diffusion, can be used to introduce the n-type impurities into the substrate through the openings provided by Mask #1. A layer of oxide, which grows during the diffusion, is then stripped. After the n^+ diffusion the wafer takes the form shown in Fig. 2C.1*b*.

An n-type epitaxial (crystalline) layer is then grown over the entire wafer. The thickness of this layer typically varies between 2 μ and 15 μ, with the thinner layers used for digital circuits and the thicker layers for analog circuits. This layer will be used for the collector region in npn transistors. The epitaxial

layer is shown in Fig. 2C.1*c*. Note that some of the impurities in the buried layer have migrated (or *out-diffused*) into the epitaxial layer during its growth. A thick layer of SiO_2 (typically 5000 Å) is then grown over the entire surface.

Mask #2 is used to pattern the SiO_2 layer for the p^+ isolation diffusion. SiO_2 is etched from the areas not photographically protected by Mask #2 to allow for this drive in diffusion following a p^+ deposition. The p^+ isolation diffusion is used to electrically separate adjacent transistors. It is wide and deep since it must completely penetrate the epitaxial layer to provide the required isolation. The wafer at this stage is as shown in Fig. 2C.1*d*. Although the isolation diffusion is shown with vertical edges in the figure, lateral diffusion, typically comparable to the vertical diffusion, causes significant out-diffusion laterally under the oxide layer, thus making the top of the channel stop considerably wider than the bottom. Following this diffusion, another thick layer of SiO_2 is grown over the entire wafer.

An optional shallow, high-resistance p-diffusion could be added at this step. This is not depicted in Fig. 2C.1 but is listed as an option in Table 2C.1. This step would provide a mechanism for making practical diffused resistors in the 1 kΩ to 20 kΩ range. A typical sheet resistance of this region would be 1 kΩ/□ to 2 kΩ/□ . Mask #A in Table 2C.1 is used to pattern these regions.

Mask #3 is used to pattern the SiO_2 layer and define the base regions for the npn transistors as well as the collector and emitter regions for lateral pnp devices. A p-type deposition and a subsequent drive in diffusion create these regions in the unprotected areas. This diffusion is much shallower than was the isolation diffusion and must not penetrate the epitaxial layer. The isolation mask openings provided by Mask #2 are typically reopened with Mask #3 to provide a few additional p-type impurities. The wafer at this stage is shown in Fig. 2C.1*e*.

Following growth of another layer of SiO_2, Mask #4 is used to pattern the emitter regions for the npn transistors. An n^+ deposition followed by a drive in diffusion creates the emitter regions. Openings are also made in the oxide above the collector to add small n^+ wells in the lightly doped collector region to provide for better electrical contact from the surface. The integrated circuit at this stage is as depicted in Fig. 2C.1*f*. The emitter diffusions must be shallow so as not to penetrate the relatively shallow p-type base regions already created. The amount and profile of the impurities in the n^+ emitter regions and the thickness of the p-type base region, which is now sandwiched between the n^+ emitter and the n-collector, strongly influence the gain of the transistor.

Mask #5 is used to pattern contact openings to allow for top contact of the circuit with the metallization. The entire circuit is then covered with a thin layer of metal. Mask #6 is used to pattern the metal, followed by the addition of a passivation layer. The completed cross-sectional view of the four components under consideration is shown in Fig. 2C.1*g*. Mask #7 patterns pad openings to allow for electrical contact to the bonding pads.

Two modifications of this process deserve mention. One involves adding a *deep collector* diffusion. This requires an additional masking step and is used to diffuse impurities under the area where the collector contacts are to be made. This step would occur either before or after the isolation diffusion and is used to

extend n^+ impurities all the way from the surface to the buried layer. Since this is such a deep diffusion, an area penalty in the collector is experienced. The deep diffusion is used to reduce collector resistance in high-current applications. The second modification involves adding an additional p-diffusion in the p-channel stops. This also requires an additional mask step and is used to avoid surface inversion in high-voltage parts. With the exception of open collector circuits, this step is not common in basic logic parts.

The npn and pnp transistors in this process are depicted in Fig. 2.2-9. The npn transistor is called a vertical npn device since the emitter, base, and collector regions are stacked vertically. It can now be seen that the n^+ buried layer decreases the collector resistance that must be modeled in series with the collector.

The pnp transistor is called a lateral device since it is stacked laterally (horizontally). The base width cannot practically be made as narrow and the base area is not as accurately controllable as for the npn device. In addition, the emitter and collector regions must have the same impurity profile. The characteristics of the lateral transistors are generally considered poorer than those of the vertical devices. Other pnp transistors, not shown in Fig. 2C.1, can be constructed by using the p-type base diffusion as the emitter, the n-type epitaxial layer as the base, and the p-type substrate as the collector. These devices are called substrate transistors. The performance of these devices is also mediocre, and applications are restricted since all collectors of substrate transistors are common.

The capacitor that was constructed in Fig. 2C.1 may at the outset appear to be merely a diode. It would serve the purpose of a diode, even though the area is considerably more than may be required in most applications. When reverse biased, however, the depletion layer forms the dielectric, and the p and n regions on either side form the capacitor plates. Capacitors made like this, with total

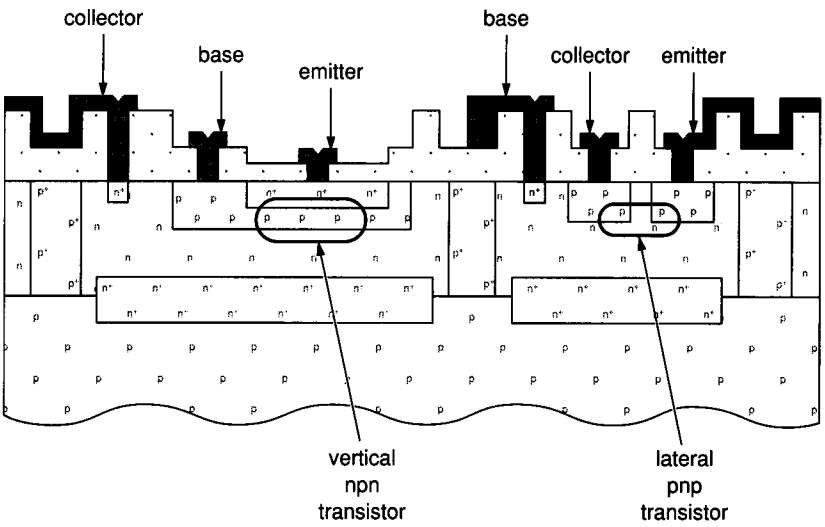


FIGURE 2.2-9
Vertical and lateral transistors in a bipolar process.

capacitances from the sub-picofarad to the 100 picofarad range, have proven practical. The capacitors are not without limitations, however. The requirement that the junction must always be reverse biased severely limits the interconnection flexibility of this device. The width of the depletion layer is voltage dependent, making the capacitance nonlinear. Temperature also affects the capacitance value. Finally, the base–emitter junction typically breaks down with a reverse bias of about 7 V, limiting the maximum voltage that can be applied to the capacitor. The base–collector junction can also be used as a capacitor. Its characteristics are very similar to the base-emitter capacitor with the exception that it offers an increased reverse breakdown at the expense of a lower capacitance density. Junction capacitors will be discussed in more detail in Chapter 3. An alternative to the junction capacitor would be a metal–oxide–semiconductor capacitor formed between the metal and the n^+ emitter diffusion. Although the characteristics of this capacitor would be better than those of the junction capacitors, an extra mask step is generally required to provide a means of selectively stripping the thick oxide above the emitter region so that a thin oxide can be regrown. The thin oxide is needed to get the capacitance density up to a practical level.

The resistor of Fig. 2C.1 is actually just a serpentined strip of the lightly doped p-type base diffusion. The underlying n-type epitaxial layer is generally contacted and taken to the highest potential in the circuit to prevent current flow into this region. Several other techniques for fabricating resistors in this process are available. They will be discussed in Chapter 3.

Although minimum feature sizes are comparable for the bipolar and MOS processes, standard bipolar processes require more area per device than do the NMOS processes. A major reason for this increased area is the deep and wide p^+ channel stops that are required for device isolation in standard bipolar processes. An alternative bipolar process using trench isolation[23] is available which offers a significant improvement in component density over the standard bipolar process.

2.2.3 Hybrid Technology

The hybrid approach to integrated circuit design involves attaching two or more integrated circuit dies (typically of different types), along with some discrete components in some cases, in a single package to form what is called a hybrid integrated circuit. It is often, and desirably, transparent to the consumer whether the circuit is monolithic or hybrid; in some cases, however, the hybrid packages are considerably larger. The hybrid integrated circuit is typically more costly than the monolithic structures. The extra cost and size of hybrid integrated circuits is offset, in some demanding applications, by improved performance capabilities.

Hybrid circuits containing discrete components occupy considerably less area than the conventional PC board/discrete component approach. They have played a major role in demanding analog signal processing applications such as high-resolution A/D and D/A converters and precision active filters. Tolerances, temperature dependence, and area-induced component value limits for resistors and capacitors in standard MOS and bipolar processes have limited the development of monolithic integrated circuits for precision continuous-time signal processing. Thick film and thin film passive components have reasonable toler-

ances, are easily trimmable, have acceptable temperature coefficients that can be tailored for tracking, and offer reasonable tradeoffs between area required and component values. These thick film and thin film networks are commonly used for the passive components in hybrid integrated circuits. A discussion of thick film and thin film processing technologies follows.

THICK FILM CIRCUITS. The thick film technology is relatively old, requires considerable area compared to monolithic circuits, can be used for relatively high-power applications, and can be applied at relatively high frequencies (up to 1 GHz) although it is typically limited to a few MHz. The increased area required by the thick film circuits is offset by the reduced cost in equipment and processing materials required for the thick film process, the latter being a small fraction of that required for either bipolar or MOS processes.

The components available in a thick film process are resistors and capacitors along with conducting interconnects. Layers of different material are successively screened onto an insulating substrate. These materials are used for resistors and conductors as well as for the dielectrics of capacitors.

The number of resistive layers varies but practical limitations generally restrict this to at most three. Typical thickness of these layers (called pastes or inks) is about 20 μ, but they may vary considerably by design. The actual thickness of these layers is not accurately controllable ($\pm$ 30%) due to limitations in the screening process itself.

The thick film process offers the most advantages for resistor fabrication. Although capacitors are often included, the electrical characteristics of thick film capacitors are not outstanding and the capacitance density is quite low. Discrete chip capacitors, which have much better characteristics than their thick film counterparts, are often bonded to thick film resistive networks in hybrid circuit applications.

The minimum conductor width in a typical thick film process is about 250 μ, and minimum resistor widths are about 1250 μ. It can be seen that these are orders of magnitude larger than the corresponding minimum feature sizes for the MOS and bipolar processes (1–5 μ).

Screening involves forcing the paste through small holes in a tightly stretched piece of fabric called a screen, typically constructed of stainless steel. The grid is quite regular. The spacing of the holes can be specified, but practical physical limitations relating to both the mechanical characteristics of the steel and the physical characteristics of the inks prevent the use of extremely fine meshes. Screens with a grid spacing ranging from 100 to 300 filaments/inch are typical. This spacing restricts thick film resolution to somewhere around 500 μ. Where paste is not desired, holes in the screen are plugged by a mask. A squeegee is used to force the ink through the unrestricted areas. Following screening, each layer is fired to harden it.

Inks are available with sheet resistances that satisfy the equation

$$1\ \Omega/\square < R_{\square} < 10\ \text{M}\Omega/\square \tag{2.2-9}$$

for fired layers 20 μ thick. This large latitude in ink characteristics allows for a wide range of resistor values. Since only one type of ink can be used for each

resistor layer, the tradeoffs between area and sheet resistance must be considered when specifying the ink sheet resistances. Even though adjusting the length is a convenient means of establishing resistor values, long thick film resistors are to be avoided because they develop "hot spots" and are difficult to trim. The "hot spots" are caused at regions where the resistive layer is a little thinner and/or narrower than surrounding regions. This causes an increased resistance in this small region, which under constant current causes increased local power dissipation. This power dissipation causes heating, which typically further increases the resistance and power dissipation. Heating causes deterioration of the film layer at these points. Deterioration in these regions can eventually result in device failure. Short, wide resistors are also to be avoided. The main problem with short, wide resistors is the inability to accurately specify the size of the resistor since the contacts will overlap with a considerable portion of the device. A reasonable rule of thumb for the allowable width (W) / length (L) ratio for a rectangular resistor is

$$\frac{1}{10} < \frac{W}{L} < 3 \tag{2.2-10}$$

Although the W/L ratio is constrained, the values for W and L remain to be specified within the design rules of the process. It is a good practice to make the resistors large if the area is available to minimize edge roughness effects, increase power dissipation capability, and make trimming easier.

Serpentined patterns such as shown in Problem 2.11 should also be avoided with thick film technology. This is due to the increased current density that will result from current crowding at the inner corners of serpentined structures.

A capacitor is constructed by screening a conductive layer, followed by a dielectric, followed by another conductor. The dielectric is generally applied in two coats to minimize pinholes, which would short the capacitor plates together. With the two layers of dielectric, the chances of a pinhole coincident with both layers are greatly reduced. Since a thick film capacitor is actually a parallel plate capacitor, the capacitance is given by

$$C = \epsilon_R \epsilon_0 \frac{A}{t} \tag{2.2-11}$$

where $\epsilon_0 = 8.854$ pF/m, A is the area of the capacitor plates, t is the dielectric thickness, and ϵ_R is the relative dielectric constant. Inks with relative dielectric constants from 10 to 1000 are available. The high dielectric constants offer a reasonable capacitance density at the expense of large and nonlinear temperature coefficients. The lower dielectric materials offer improved performance but are restricted to applications requiring small capacitors due to a low capacitance density. As in the MOS and bipolar processes, the upper plate of thick film capacitors is typically a little smaller than the lower to minimize capacitance changes caused by minor misalignments. The two conductive layers used for the capacitor plates also serve as interconnects. If crossovers of two conductors are required, the dielectric layer can be used as an insulator at the expense of creating a small parasitic capacitor at the crossover.

The screen geometries, along with a cross-sectional view of a typical thick film process, are depicted in Fig. 2D.1 of Appendix 2D. This process has two

resistive screenings as well as two conductive layers and a dielectric for capacitor fabrication. The major process steps are listed in Table 2D.1. Process parameters and characteristics, along with design rules for a typical thick film process, are also given in Appendix 2D.

Thick film resistors can be trimmed with a laser or by abrasion. These trims, which can be very accurate, remove part of the thick film layer and thus increase the resistance. For this reason, resistors that are to be trimmed are typically targeted to be undersized in value by about 40% to guarantee trimmability in spite of process variations.

Thick film capacitors can also be trimmed by abrasively removing part of the upper plate (along with some dielectric). Since this decreases the capacitance, the thick film capacitors are typically targeted to be oversized by about 40%.

THIN FILM CIRCUITS. The components available in thin film processes are resistors and capacitors, although often only resistors are included due to both the specific applications which naturally benefit from thin film technology and the practical limitations of thin film capacitors. Thin film circuits are much smaller than thick film circuits. They are similar to thick film circuits in that successive layers are applied to an insulating substrate as contrasted to the MOS and bipolar processes, where some of the processing steps involve diffusions that actually penetrate the substrate. For conductors, thin film thicknesses are typically from 100 to 500 Å although thicknesses of several thousand angstroms are occasionally used if a high conductivity is needed. Film thicknesses from 100 to 2000 Å for resistors and film thicknesses in the 3000 Å region for dielectrics of capacitors are common. Note that these film thicknesses are comparable to the thicknesses of layers applied in the MOS and bipolar processes but are orders of magnitude thinner than the 20 μ (200,000 Å) typical of thick films. The sheet resistance range for thin film resistors is typically from 50 Ω/□ to 250 Ω/□, which is considerably less than is available in thick film processes. The thin film layers are applied by uniformly coating the entire wafer with the film. Then unwanted areas are selectively patterned and etched with a photolithographic process similar to that used in the MOS and bipolar cases.

The minimum feature sizes for the thin film components are comparable to those of the MOS and bipolar processes. The temperature characteristics and performance of thin film components are quite good with the exception of the dielectrics for capacitors, which are quite lossy. "Hot spots," which were a problem with thick film circuits for long resistors, are not a major problem with thin film resistors since the thin films are typically more uniform and since thin film applications generally require smaller current flow.

Thin film circuits are much more expensive to produce than their thick film counterparts because of the sophisticated equipment that is needed for both the photolithographic process and the film depositions and etching. They are used extensively in telecommunication circuits at low frequencies but also find applications at higher frequencies (up to 30 GHz) as well.

Thin film resistors can be accurately trimmed by a laser. Thin film capacitors are not well adapted to a continuous trim although binarily weighted capacitors connected with laser-fusible links are trimmable in quantized decrements.

Thin films can also be applied on top of monolithic structures, offering considerably more performance capability than is attainable with either the thin film or monolithic approaches themselves. Problems with film technology and the decreased yield per wafer associated with both increased die area and an increased number of processing steps have slowed the development of such processes.

The mask layouts, along with cross-sectional views of a typical thin film resistive process, are depicted in Fig. 2E.1 of Appendix 2E. This is a subtractive process because all film layers are added prior to any patterning, and it has a single resistive layer and a single conductive layer. The major process steps are listed in Table 2E.1 of the same appendix. Process parameters and design rules for a typical thin film process are also given in Appendix 2E.

2.3 DESIGN RULES AND PROCESS PARAMETERS

Design rules are generally well-documented specifications listing minimum widths of features (conductor, moat, resistor, etc.), minimum spacings allowable between adjacent features, overlap requirements, and other measurements that are compatible with a given process. Factors such as mask alignment, mask nonlinearities, wafer warping, out-diffusion (lateral diffusion), oxide growth profile, lateral etch undercutting, and optical resolution and their relationships with performance and yield are considered when specifying the design rules for a process. It is not our intention in this text to rigorously investigate the technical details about how design rules are derived but rather to consider the design rules and process parameters as a set of constraints within which the circuit designer must work. This is justifiable because the basic format of the design rules and process parameters remains relatively fixed, with changes in the process contributing only to numerical perturbations of the design rules and process parameters. The design rules, the process parameters, and their relationship with device characteristics serve as an interface between the process engineers and the circuit designers. Both groups, along with representatives from marketing (since yield is affected by the design rules), have input into the evolution of these interfaces.

Although the minimum feature sizes, which ultimately determine the design rules, have been steadily decreasing with time to the benefit of yield and production costs, it is important that designers adhere to the design rules once a process has been selected for a particular project. Most large semiconductor houses have developed or purchased sophisticated computer software to verify that layouts violate no design rules. In the process of verifying design rules, it is often the case that layout errors are also detected since these errors will often violate a design rule. For large designs that involve thousands of transistors, it is crucial that these verifications be made since a single design rule violation or layout error will often be fatal (i.e., the circuit won't work). The importance of adhering to design rules and utilizing verification programs for simple as well as complicated designs cannot be overemphasized. If design rules are intentionally violated, the verification software cannot be fully utilized and perhaps not utilized at all.

Typical sets of design rules and process parameters for the NMOS, CMOS, bipolar, thick film, and thin film processes discussed in Sec. 2.2 are summarized

in Appendices 2A–2E. Some of the parameters listed in those appendices have not been defined yet but will be discussed in Chapter 3. The 3 μ NMOS parameters and the 3 μ CMOS parameters are very similar to those provided by MOSIS for their 3 μ NMOS and 3 μ CMOS processes in 1988. We have attempted to maintain most of the notation established by MOSIS to aid students who will be doing designs in a MOSIS process.

A discussion of selected key design rules and process parameters for the CMOS process follows. Many of these comments apply either directly or with obvious modifications to the other processes. In the interest of conserving space, the interpretation of the parameters and design rules for the remaining processes will be left to the reader.

CMOS DESIGN RULES. The CMOS design rules are listed in Table 2B.2 of Appendix 2B. These rules are depicted graphically in Table 2B.3.

The design rules list guidelines (actually restrictions) about how each of the geometrical figures on each mask level align relative to each other and to other mask levels. Unless specifically stated to the contrary with the comment "exactly," all rules are minimum spacings between the corresponding geometrical figures. In general, the designer may exceed these minimum spacings to whatever degree is deemed appropriate. It should be emphasized, however, that from both cost and performance viewpoints, the die area should be as small as is practical. At this stage one might be tempted to conclude that a margin of safety, or improved reliability, could be obtained if a more conservative set of rules were established by the designer. With the exception of some matching considerations that will be discussed later in this book, few benefits are derived from following this strategy since an economically motivated margin of safety was considered when establishing the design rules. The rules were derived under the assumption that large circuits with many devices sized at the minimum allowable levels must have good performance and high yield.

In Table 2B.2, two sets of dimensions are specified. The first corresponds to those specified in a 3 μ CMOS process provided by MOSIS. The second set is in terms of the scaling parameter, λ, which characterizes the feature size of the process. The feature size (minimum poly width, active width, and metal width) is 2λ. This parameterization is used so that the design rules do not need to be rewritten as the feature size of the process shrinks. Substituting $\lambda = 1.5$ μ will give a process very similar to the 3 μ process characterized in the first column of this table. Although similar in intent, the scalable parameters listed in this table are not identical to those characterizing the MOSIS scalable CMOS process.

The mask geometries themselves may differ somewhat from the geometries specified for the corresponding geometrical feature by the designer. The exact geometries specified by the designer are termed *drawn* features. The change in feature sizes on a mask from those specified by the designer is termed *size adjust* and is undertaken in cooperation with those responsible for the processing. This allows for precompensation of effects, such as lateral etching or out-diffusion, that make the physical dimensions on silicon different from the mask dimensions. Size adjust is often used so that the "effective" dimension (the physical dimension

realized after fabrication) is nominally equal to the drawn dimension. Specific comments about selected key rules follow.

p-well. The spacing (Rule 1.2) between two p-wells at different potential is very large. This allows for accommodation of the lateral diffusion. Since the p-well is very deep (3–4 μ), the lateral diffusion is also significant.

Via. The via level is used for interconnecting Metal 1 and Metal 2. The via design rules are very similar to those for the contact openings. Both involve making openings in thick oxide layers.

Pads. The dimensions for the metal bonding pads have not scaled with decreases in feature size of the processes. The size of the capillaries used for attaching bonding wires, which are several mils in diameter to accommodate the nominal 1 mil bonding wire, has not decreased significantly for a long while, thus necessitating both large bonding pads and large spacing. Smaller probe pads are often included on designs to facilitate diagnostic probing during the debugging stage.

Active. Size adjust, sometimes termed mask bias, is used to preadjust feature sizes on the active mask so that the targeted effective feature sizes are close to the drawn feature sizes. This allows for compensation of the field encroachment into the active region. Also, this size adjust thus compensates for the width reduction experienced during processing. Note that a large spacing between active and p-well is provided to accommodate for the lateral p-well diffusion.

Poly. The poly overlap rules are primarily to provide compensation for mask alignment errors between poly and active. Although the process is self-aligned, this self-alignment is achieved only for modest misalignment of devices. If, for example, the misalignment is so bad that the poly does not entirely cover the active region, then the subsequent n^+ or p^+ drain and source diffusions will create a conductive region between the drain and source that cannot be controlled by the gate. Design Rule 3.4 governs this concern. Size adjust is used on the POLY mask to do length adjustment for lateral diffusion.

If a second poly layer is available, it may not cross a boundary of POLY I. The sharp edge of POLY I would make the step coverage of this edge with the thin oxide unreliable. The integrity of the POLY II over this step would also be in question. Nor is POLY II permitted for transistor gates although such devices would likely be functional. The main purpose of POLY II is as a second plate on poly–poly capacitors although some applications as an interconnect medium or as a resistor may exist.

p^+ select. The p^+ select rules are used primarily to allow for mask misalignments with the active mask and the poly mask.

Contact. The contact openings are specified exactly rather than minimally. Although it is often the case that a large contact between two regions is desirable to reduce contact resistance, a large single contact opening is not permitted; rather,

numerous separate contacts must be used. Multiple individual contacts were used for contacting the p-well in Fig. 2B.1*a*. The reason for restricting the contact opening size and making all essentially the same can be best appreciated by considering what the contact-opening etch must accomplish. The contact openings must penetrate a thick (5000 to 7000 Å) layer of oxide. Thus, this step must consume much more oxide than is required during the thin oxide stripping steps. If a large contact opening were permitted, the central areas of these openings would be etched away before the oxide was completely removed for the smaller contact openings. As the etching continued in the small openings, any pinholes in the large open area could be further attacked by the etchant. Since the underlying poly layer thicknesses and diffusion depths are comparable in thickness to the layer that must be removed during contact openings, and since underlying thin oxides are *much* thinner, these pinholes could cause device failure. Although the probability of failure due to a single larger contact may be very low, the probability of a single failure that would render a circuit defective if a large number of these large openings were permitted may be unacceptably large. Even if the pinholes did not cause shorting, the reliability of such devices deteriorates with increased risks of premature device failures after the part is in use. For these reasons, contact openings on the gates of transistors (no contact to poly inside active) are usually not permitted either.

Contact openings to POLY II on top of POLY I are permitted. Although this type of contact is also plagued by the pinhole problem, the POLY II layer is generally used in analog applications as an upper plate of a capacitor. The total number of capacitors in these circuits is generally quite small compared to the number of transistors in a large digital circuit, thus minimizing (in the probabilistic sense) the failures due to the pinhole problem (see Problem 2.12).

Example 2.3-1. A thick film resistor layout strategy is shown in Fig. 2.3-1. Determine the minimum substrate size ($L \cdot H$) if the resistor is to have a nominal value of 325 kΩ and is to use the layout strategy of this figure. Use the design rules of Table 2D.2 ($\lambda = 250\ \mu$) and the process parameters of Table 2D.3.

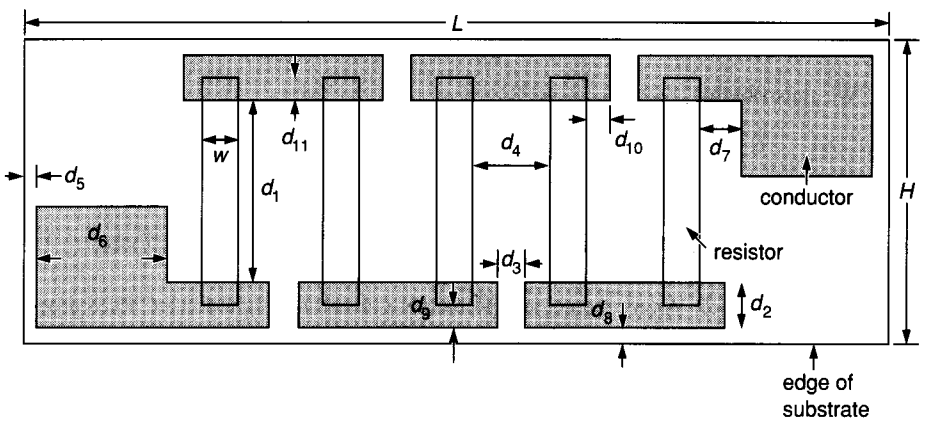


FIGURE 2.3-1
Layout for Example 2.3-1.

Solution. Since $R = 325\ \text{k}\Omega$ and since the goal is to minimize area, the highest sheet resistance screening will be used for the resistors. From Table 2D.3, we obtain $R_\square = 10{,}000\ \Omega/\square$. If the contact resistance and sheet resistance are neglected, we require $n = 325\ \text{k}\Omega/R_\square = 32.5$ squares of resistive material. Since there are five equal width resistors in series, each of width W, the length of each (d_1) must equal $32.5W/5 = 6.5W$.

From Fig. 2.3-1, it follows that the length and height of the substrate are given, respectively, by

$$L = 2d_5 + 2d_6 + 2d_7 + 5W + 4d_4$$

$$H = 2d_8 + 2d_2 + d_1$$

From Table 2D.2, the following minimum sizes are obtained for each of the parameters d_2–d_{11} and W.

Parameter	Rule	Size
d_2	1.1	2λ
d_3	1.2	2λ
d_4	2.2	2λ
d_5, d_8	1.5	2λ
d_6	4.0	4λ
d_7	2.3	2λ
d_9	1.3a	λ
d_{10}	1.3b	λ
d_{11}	2.4	λ
d_1, W	2.1	5λ

It should be noted that even though these are minimum spacings, the layout itself may impose more stringent requirements to avoid layout violations. Specifically,

$$d_4 \geq d_3 + 2d_{10} = 4\lambda$$

$$d_2 \geq d_9 + d_{11} = 2\lambda$$

$$d_1 \geq 5\lambda$$

We thus obtain

$$L = (2)(2\lambda) + (2)(4\lambda) + 2(2\lambda) + (5)(5\lambda) + (4)(4\lambda) = 57\lambda$$

$$H = (2)(2\lambda) + (2)(2\lambda) + (6.5)(5\lambda) = 40.5\lambda$$

With $\lambda = 250\ \mu$, the minimum substrate size is 1.425 cm × 1.0125 cm.

Example 2.3-2. An n-channel and p-channel transistor designed in a CMOS process are depicted in the layout of Fig. 2.3-2. If the p-channel device is of size $L/W = 6$ and the n-channel device is of size $L/W = \frac{1}{2}$, determine the minimum distance between the edge of the p-tub and the edge of the p-moat, d, shown in the figure if the devices are packed as closely as possible following the layout placement shown. Assume the devices are designed in the CMOS process discussed in Sec. 2.2 with the design rules of Table 2B.2. Assume $\lambda = 0.75\ \mu$.

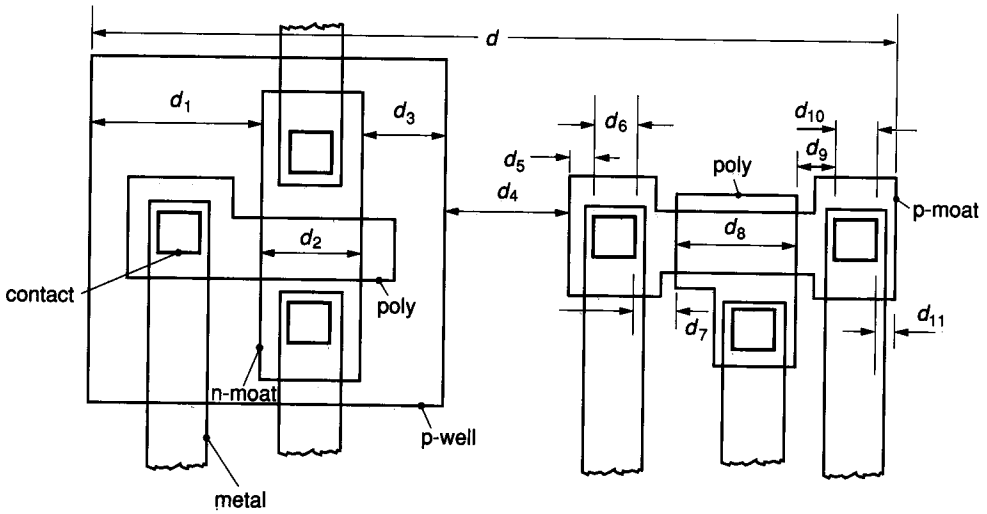


FIGURE 2.3-2
Layout for Example 2.3-2.

Solution. Since only the L/W ratios are specified, W will be chosen minimum size for the n-channel transistor and L will be chosen minimum size for the p-channel device.

The dimensions d_1–d_{11} must satisfy the design rules of Table 2B.2. The restriction on these dimensions as imposed by the design rules are as follows.

Dimension	Design rule	Minimum size
d_1, d_3	2.5	2λ
d_2, d_8	2.1, 3.1	2λ
d_4	2.3	6λ
d_5, d_{11}	5.8	λ
d_6, d_{10}	5.1	2λ
d_7, d_9	5.6	2λ

The layout itself introduces some additional constraints. From the L/W ratios specified, it can be concluded that $d_2 \geq 2L_{\min}$ and $d_8 \geq 6W_{\min}$. From design rules 2.1 and 3.1 of Table 2B.2, it can be concluded that $d_2 \geq 4\lambda$ and $d_8 \geq 12\lambda$. The layout itself also places constraints on d_1. Since the contact as well as the edges of the metal and poly on the left-hand side of the n-channel transistor are to remain inside the p-well (a somewhat arbitrary requirement) it can be argued by design rule 5.6 that the contact must be at least 2λ from the n-moat, and from design rule 5.1 that the contact width must be 2λ. From design rules 5.4 and 5.7, the poly and metal must overhang the contact by λ. Since there is no minimum spacing between the poly and p-tub, the left-hand edge of the poly in the tub can be coincident with the tub edge (although the guard ring may be broken). It thus follows that

$$d_1 \geq 5\lambda$$

But

$$d = \sum_{i=1}^{11} d_i$$

which from above becomes d $= 39\lambda = 29.25\ \mu$.

2.4 LAYOUT TECHNIQUES AND PRACTICAL CONSIDERATIONS

Once a circuit design is complete, it becomes necessary to provide an area-efficient layout of the circuit to generate the masks necessary for fabrication. Although at the outset it appears that the circuit designer's job is complete at this point and that the layout can be undertaken by a draftsman (as is commonly done with PC board versions of discrete component designs), this is far from the case in IC design. Some companies provide draftsmen for this purpose who interact closely with the designers, whereas other philosophies leave this task entirely to the design engineer. In either case, the design engineer is still involved at this stage because component sizing and spacing, as well as the parasitics associated with integrated circuit components, must typically be considered in the design itself since their effects are often significant. This is particularly important in designs including analog circuitry. Even the opening sentence of this paragraph is an oversimplification of the situation since the initial design itself will likely not be complete until the layout is finished—as was indicated in Fig. 1.3-1. Although the design engineer is typically not responsible for any steps in the fabrication process once the mask information has been delivered, he or she is generally still responsible for the project until the product is in production. For complicated circuits, first silicon (the physical integrated circuits produced by the initial design) will generally not be acceptable for marketing because of either the circuit's failure to meet some specifications or total circuit failure caused by (1) a design or layout error, (2) failure of the designer to adequately account for all relevant parasitics, or (3) unacceptably low yield due to failure to center the design parameters appropriately in the actual process window.

One of the first considerations in the layout is sizing the devices as well as the interconnections. For the long rectangular resistor shown in Fig. 2.4-1, the total resistance is obtained from the sheet resistance by the expression

$$R = (L/W)R_{\square} \tag{2.4-1}$$

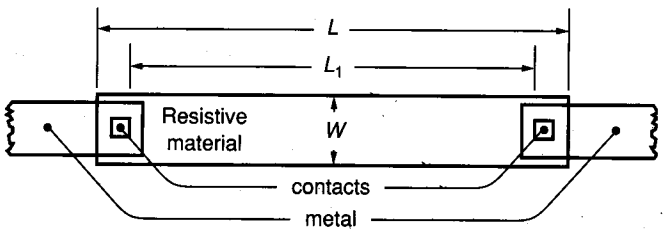


FIGURE 2.4-1
Resistor component sizing considerations.

provided that L is long enough so that the difference between L and L_1 is negligible (because $L - L_1$ is fixed by the design rules). Since the L/W ratio rather than either L or W determines the resistance, it may seem to make little difference what the values of L and W actually are, but this is often not the case. In addition to the difficulty in determining the effective length (L or L_1?) for short resistors, the edges will typically be somewhat rough due to unevenness in processing. This unevenness will cause variations in resistance. Making W and L larger reduces the relative effects of both the edge variations and the difference between L and L_1. In addition, the power-handling capability will be increased with larger devices. These improvements with larger devices are obtained at the expense of increased area, which will reduce the number of dies per wafer.

For MOSFETs, it will be shown in Chapter 3 that the length/width ratio, rather than the length and width themselves, plays one of the major roles in the device model. Again, increasing the area for a fixed length/width ratio will reduce the effects of unevenness in the edges, but this is obtained at the expense of increased area and increased gate capacitance.

As mentioned earlier, it is common practice when laying out capacitors to make one plate (typically the upper) a little smaller than the other so that the smaller plate effectively defines the plate area even if minor misalignments in the masks occur. This also makes the edge field effects easier to account for. Although the area of the smaller plate essentially determines the capacitance, the geometry of this plate itself remains to be determined. If the relationship between this capacitor value and others in the circuit is not of major importance, the shape of the capacitor will likely be selected to conform to available spaces around adjacent components, thus minimizing circuit area.

Many applications require that resistor or capacitor ratios be accurately determined. This is particularly common in analog signal-processing circuits. Ratio matching requirements of 1% to 0.1% or better are common. Although absolute component value tolerances better than 1% (or even 10%) are not currently feasible without trimming in any of the processes discussed in Sec. 2.2, the ratio accuracy specified above is attainable in some processes and is maintainable over a wide range of temperature. For resistor layouts both the individual L/W ratios, as well as the area and shape, become design parameters available to the circuit designer. The area of the resistors should be large enough to make the effects of edge roughness acceptably small, but they should be small enough to make the circuit economical and to avoid deviations caused by global variations in processing characteristics.

An example of realizing a resistor with a 3:1 ratio to R1 by three different techniques is shown in Fig. 2.4-2. Since conductors are quite good, R4 offers several distinct advantages over R2 and R3 for attaining this ratio. Comments about the different approaches follow.

1. The long resistor, R2, often cannot be conveniently placed on the circuit in an area-efficient manner. Furthermore, the question of exact length remains open and the fact that the number of contacts are not related by the 3:1 ratio limits the accuracy of R2/R1.

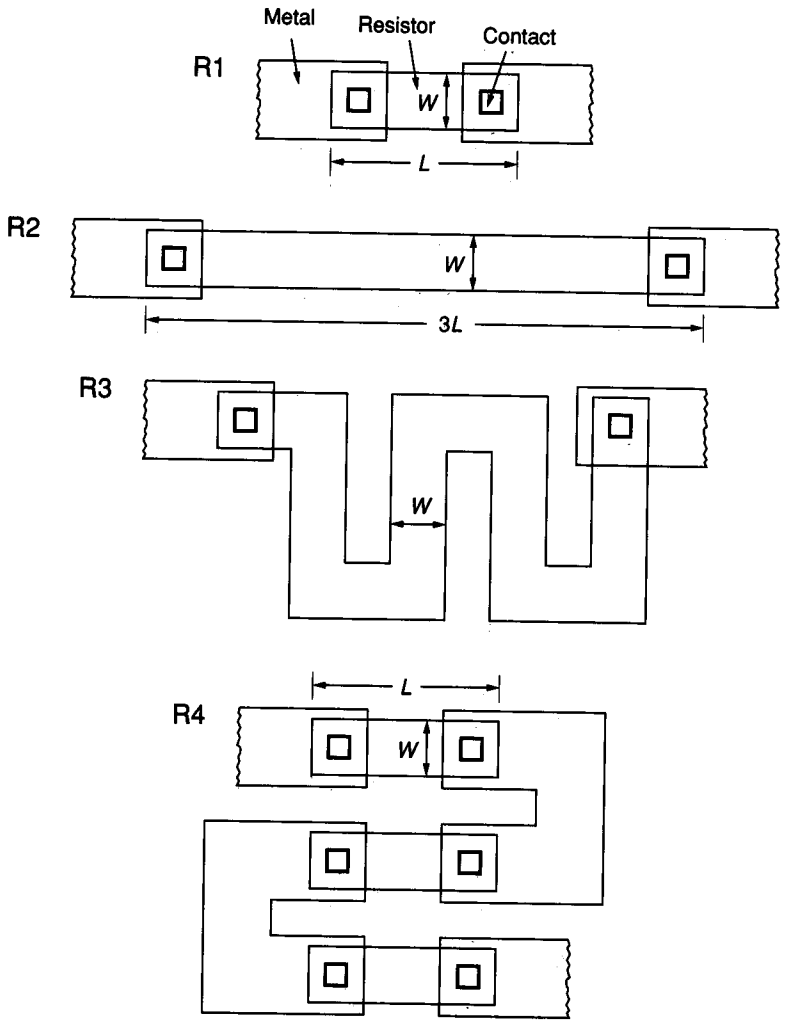


FIGURE 2.4-2
Resistor ratio-matching considerations.

2. The serpentine pattern used for R3 is quite common to keep overall aspect ratios practical. However, the difficulty in accurately accounting for the corners (using the .55 rule) and the differences in periphery length will generally make the R3/R1 ratio the least accurate of the schemes shown in the figure. The contact resistances are also not accounted for in the ratio with R1.
3. Even though the exact "length" is difficult to define, the three serpentined resistors in R4 are ideally identical to R1, so the ratio accuracy is maintained. This approach also accounts for any contact resistance associated with the contacts themselves as well as differences in temperature characteristics of the resistive and contact regions.

For realizing capacitor ratios, the area should be large enough to make the effects of edge roughness acceptably small. If the areas are too large, however, the circuits become impractical both because of the area requirements and because of the increased failure rate due to an increased likelihood of dielectric defects (one or more pinholes), which will short the capacitor plates together. Variations in dielectric thickness must also be considered if large areas are involved. In addition to maintaining the required ratio, the periphery lengths should also adhere to the ratio if possible. Fig. 2.4-3 shows three different methods of realizing a 3:1 capacitance ratio to C1. Several comments about these approaches follow.

1. C2 maintains the same geometry but the length of the perimeter differs somewhat. This will limit ratio accuracy due to variations in etching of the POLY

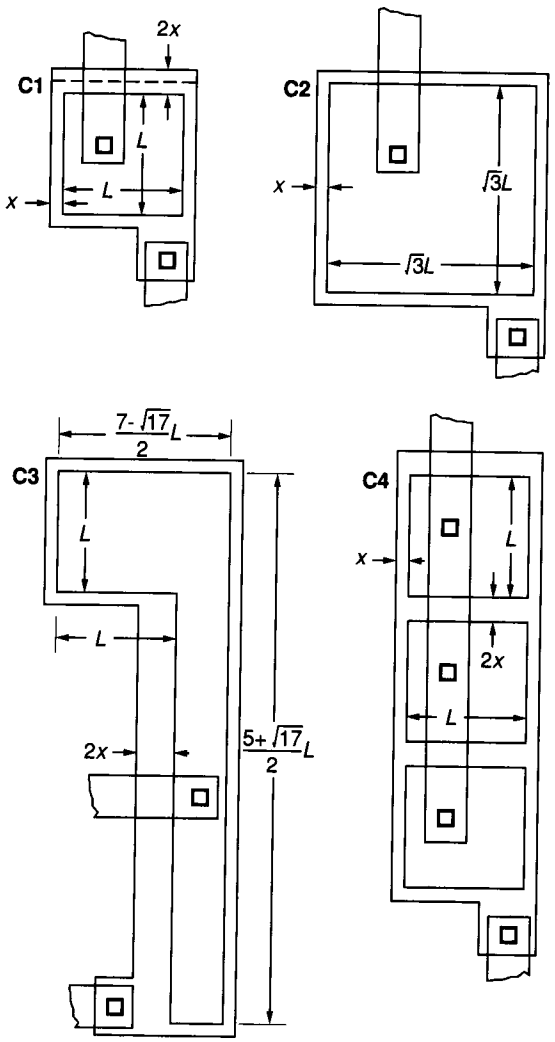


FIGURE 2.4-3
Capacitor ratio-matching considerations.

II edge. Furthermore, the ratio of the small capacitance from the conductor to the lower plate of C2 is 1:1 instead of 3:1.

2. C3 has a periphery that is three times that of C1, but the small capacitance from the conductor to the lower plate is still 1:1 with that of C1. The number of inside and outside corners in C1 and C3 does not ratio by 3:1, thus further limiting accuracy.
3. C4 has the same periphery as C1 as well as the same parasitics from the conductor to lower plates. This will give the best ratio of the three approaches. If the capacitor areas are large, even this approach will be affected by variations in dielectric thickness. Since the oxide layer is typically quite uniform locally, the ratio accuracy for large areas can be improved if each of the capacitors is further subdivided in an identical manner and the smaller capacitors for C1 and C4 are interleaved.

Example 2.4-1. Assume the four capacitors of Fig. 2.4-3 are fabricated with a dielectric between two layers of polysilicon and that the design value of L is 25 μ. Assuming the POLY I–POLY II capacitance density is 0.7 fF/μ^2, the POLY I–metal capacitance is .04 fF/μ^2, the nominal Poly I overlap of Poly II (x in Fig. 2.4-3) is 2 μ and all metal widths are 3 μ. Determine the ratios C_2/C_1, C_3/C_1 and C_4/C_1 if overetching of the upper plate (which was not accounted for during design) during etching was 0.5 μ, and calculate the effective number of bits of ratio-matching accuracy in each case.

Solution. With $L = 25\,\mu$, the area of the upper plate of C1 is $A_{1P} = [25 - 2(0.5)]^2$ $\mu^2 = 576\,\mu^2$, and the area of the metal overlap of Poly I is $A_{1M} = (3\,\mu)(4\,\mu + 0.5\,\mu) = 13.5\,\mu^2$. Thus, $C_1 = (A_{1P})(0.7\text{ fF}/\mu^2) + (A_{1M})(0.04\text{ fF}/\mu^2) = 403.74$ fF. Likewise, $C_2 = 1252.878$ fF, $C_3 = 1208.98$ fF, and $C_4 = 1211.22$ fF. Thus, $C_2/C_1 = 3.103$, $C_3/C_1 = 2.9945$, and $C_4/C_1 = 3.0$. To determine the effective number of bits of resolution in each case, assume that the effective number of bits, n, is the maximum integer value of n that satisfies the expression

$$2^n < \frac{\text{nominal ratio}}{|\text{nominal ratio} - \text{actual ratio}|}$$

Thus, the effective number of bits of C_2/C_1 is 4, and of C_3/C_1 is 9. Based upon the model presented here, the C_4/C_1 ratio is perfect. Edge roughness and oxide thickness variations will practically limit the effective number of bits in the C_4/C_1 ratio to the 8–11 bit range with existing processes. From this example, it should be apparent that care in layout can have a major impact on ratio-matching accuracy.

Parasitics can cause significant deviations in circuit performance and should be minimized during layout. Some of the common parasitics encountered are (1) the resistances associated with polysilicon and doped semiconductor regions when used as conductors, and (2) the capacitances associated with any crossover, from any conductor to substrate, and with any depletion region in a reverse-biased pn junction. Unfortunately, these resistive and capacitive parasitics can be comparable in magnitude to the desired component values to which they are connected if good layout rules are not established. Even with good layout rules,

the values of these parasitics may be significant. Clever design techniques and inclusion of the unavoidable parasitics in the analysis when possible, however, help overcome some of the parasitic limitations. The parasitics associated with a depletion region of a reverse-biased pn junction are particularly troublesome since they are voltage dependent and thus difficult to properly account for in analysis and design. Even if accounted for, the parasitic capacitances often cause unwanted *cross talk* between signal paths, and the nonlinear capacitors can cause nonlinear signal distortion that may be unacceptably large.

Cleverness in layout will also often save a considerable amount of area. Even though the concern for minimizing area is always present, it is especially important to minimize area in small digital blocks that will be repeated thousands of times in high-volume VLSI circuits. Three different techniques for connecting the gate to the drain of an enhancement MOSFET are depicted in Fig. 2.4-4. The layout of Fig. 2.4-4*a*, which uses a conventional metal interconnect, requires considerable area. The circuit of Fig. 2.4-4*b* is more area-efficient, but it is not allowed in many processes because of the concern of reduced yield associated with having pinholes, which cause device failure when gate contact is made in the channel. The connection of Fig. 2.4-4*c*, which is termed a *butting contact*, is the most area-efficient, although butting contacts are only available in some processes. Note that a single contact opening is used for the butting contact.

A doped semiconductor region, a second conductor, or a polysilicon strip can be used as a conductor or as a crossover, provided that an insulating layer exists between the devices. Crossovers are often required since jumper wires are totally impractical in integrated circuit design.

In the double-poly NMOS process described in Sec. 2.2, three-level conductor stacking (metal, POLY I, and POLY II) is possible although only a two-level crossover of Poly I and metal is permitted since Poly II cannot cross a Poly I boundary. Metal over a moat diffusion will also serve as a crossover. In the CMOS process previously described, poly–metal, metal–p-well, and metal–moat (either n^+ or p^+) overlaps can all serve as two-level crossovers. In the bipolar process described metal–n^+ or metal–p (base diffusion) overlaps make

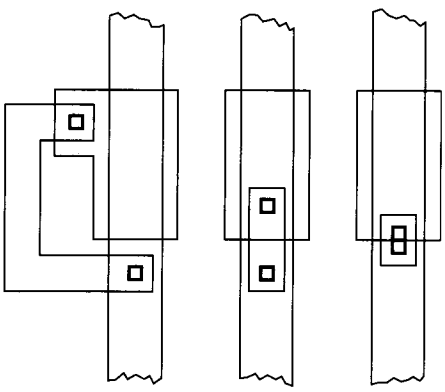


FIGURE 2.4-4
Poly-to-moat contacts: (*a*) Contact off of channel, (*b*) Contacts above channel, (*c*) Butting contacts.

practical two-level crossovers. A metal–epitaxial crossover is also possible, but the resistance associated with the epitaxial layer may be too high to make this practical in many applications.

Most design groups do the layout on an interactive graphics computer system, which serves as a workstation. As capability increases and price decreases, the trend is toward providing each design engineer with a dedicated workstation. Any or all mask levels can be displayed simultaneously, and rapid zooms and scrolls are generally available to provide both local and global information. Once the individual polygons or macros characterizing a mask feature are placed on the CRT, they are added to the geometrical database. After the layout in the CRT is in its final form, so is the database.

As an alternative to the engineer's part in layout, considerable research effort in the past 10 years has been devoted to automatic layout and routing programs. With this approach the computer will place the components and provide all appropriate interconnects (route) directly from the electrical description of the circuit. In addition to saving the layout time, layout errors can be eliminated with these programs. Programs that perform this task are available today and are quite useful for low-volume products requiring a fast turnaround where the area of the IC itself is not of major concern. Widespread use of automatic routing programs is particularly apparent with gate array products, although some of these programs get "stuck" in complicated regions and need human interaction to overcome these problems. Handpacked designs are generally denser than those obtainable with the automatic layout and routing programs. The challenges associated with the automatic layout and routing problem were recognized years ago from research on PC board layout.

Interactive layout and routing routines are also useful since the computer can rapidly do the drafting required in nonchallenging portions of the circuit, whereas the operator can often make better choices for placement or routing in tight situations. Although it may appear to the unwary that the development of such a program is straightforward, the challenge of such a program is attested to by the fact that many millions of dollars have been spent on the development of the routing and placement programs over the past decade, and research activity in this area is still intense.

Layout verification programs, such as Design Rule Checkers (DRCs), are useful for confirming that no design rule violations occur and for detecting some layout errors (since these errors often cause a design rule violation). For large circuits, considerable amounts of computer time are required for most sophisticated verification programs. Some layout editors incorporate the DRC algorithms into the layout program and do local design rule checks each time a feature is added to the database, thus ensuring that the database violates no design rule as it is created. One of the more popular programs in this class is MAGIC from the University of California at Berkeley. Automated schematic extraction or hand generation of a circuit schematic by someone other than the initial designer (who may be too familiar with the design and thus more likely to pass over an error again), followed by comparison with the original schematic, is often useful for detecting layout errors.

REFERENCES

1. N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design: A Systems Perspective*, Addison-Wesley, Reading, MA, 1985.
2. W. S. Ruska, *Microelectronic Processing*, McGraw-Hill, New York, 1987.
3. A. B. Glaser and G. E. Subak–Sharpe, *Integrated Circuit Engineering*, Addison-Wesley, Reading, MA, 1977.
4. S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed., Wiley, New York, 1981.
5. A. B. Grebene, *Bipolar and MOS Analog Integrated Circuit Design*, Wiley, New York, 1984.
6. S. K. Ghandi, *VLSI Fabrication Principles*, Wiley, New York, 1983.
7. "MOSIS User's Manual," USC Information Sciences Institute, Marina del Rey, CA.
8. A. Makherjee, *Introduction to NMOS & CMOS VLSI System Design*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
9. D. A. Pucknell and K. Eshraghian, *Basic VLSI Design Principles and Applications*, Prentice-Hall, Sydney, Australia, 1985.
10. W. C. Till and J. T. Luxon, *Integrated Circuits: Materials, Devices, and Fabrication*, Prentice-Hall, Englewood Cliffs, NJ, 1982.
11. W. Maly, *Atlas of IC Technologies: An Introduction to VLSI Processes*, Benjamin/Cummings, Menlo Park, CA, 1987.
12. A. Sedra and K. Smith, *Microelectronic Circuits, 2nd ed.*, Holt, Rinehart, and Winston, New York, 1987.
13. P. Chirlian, *Analysis and Design of Integrated Electronic Circuits*, Harper and Row, New York, 1987.
14. D. Hodges and H. Jackson, "Analysis and Design of Digital Integrated Circuits," McGraw-Hill, New York, 1983.
15. J. Millman, *Microelectronics, Digital and Analog Circuits and Systems*, McGraw-Hill, New York, 1979.
16. P. Gray and R. Meyer, *Analysis and Design of Analog Integrated Circuits, 1st or 2nd ed.*, Wiley, New York, 1977 or 1984.
17. J. Millman and A. Grabel, *Microelectronics*, McGraw-Hill, New York, 1987.
18. E. A. Vittoz, "The Design of High–Performance Analog Circuits on Digital CMOS Chips," *IEEE J. Solid State Circuits*, Vol. SC-20, pp. 656–665, June 1985.
19. J. B. Shyu, G. C. Temes, and F. Krummenacher, "Random Error Effects in Matched MOS Capacitors and Current Sources," *IEEE J. Solid State Circuits*, Vol. SC-19, pp. 948–955, Dec. 1984.
20. J. L. McCreary and P. R. Gray, "All-MOS Charge Redistribution Analog-to-Digital Conversion Techniques—Part I," *IEEE J. Solid State Circuits*, Vol. SC-10, pp. 371–379, Dec. 1975.
21. Y. S. Lee, L. M. Terman, and L. G. Heller, "A 1 mV MOS Comparator," *IEEE J. Solid State Circuits*, Vol. SC-13, pp. 294–298, June 1978.
22. J. L. McCreary, "Matching Properties, and Voltage and Temperature Dependence of MOS Capacitors," *IEEE J. Solid State Circuits*, Vol. SC-16, pp. 608–616, Dec. 1981.
23. A. H. Shah, et al, "A 4-MBit DRAM with Trench-Transistor Cell," in *IEEE J. Solid State Circuits*, vol. SC-21 (1988), pp. 618–626.

PROBLEMS

Section 2.2

2.1. Sketch a cross-sectional view of the circuit of Fig. 2A.1 in Appendix 2A after fabrication is complete along section line CC′ of Fig. 2A.1*a*. Repeat with section line CC′, DD′, EE′, and FF′ of Fig. 2B.1*a* and CC′ of Fig. 2C.1*a*.

2.2. A POLY I resistor has a TCR of 0.1%/°C. What % change in R will be obtained if the temperature of the resistor changes from 0° to 100°C?

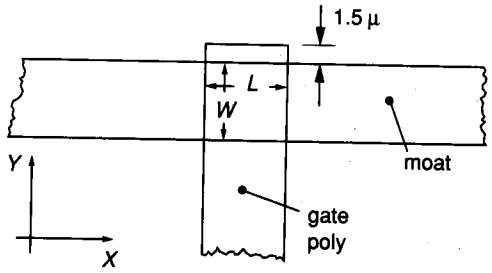


FIGURE P2.5
Problem 2.5.

2.3. What is the maximum permissible current in a minimum width metal trace (Metal I) in the CMOS process of Appendix B if metal migration is to be prevented?

2.4. Compare the incremental cost (cost of adding an additional device) of fabricating a 3 $\mu \times 3\ \mu$ transistor, a 1 pF capacitor, and a 10 kΩ resistor designed in the process of Appendix 2B assuming the typical processing costs of Table 1.4-2.

2.5. If the poly gate on the device shown in Fig. P2.5, which ideally has $W = 2.5\ \mu$ and $L = 2.5\ \mu$, has a $\pm 1.25\ \mu$ variation in either the X or Y direction due to registration limitations, what is the worst case W/L ratio due to improper registration?

2.6. It is desired to monolithically realize the first-order lowpass filter shown in Fig. P2.6 with a 3 db cutoff frequency of 1 kHz [i.e., $RC = 1/(2\pi 1000)$]. If the resistance density is $R_d = 1.6\ \Omega/\mu^2$ and the capacitance density is $C_d = 0.7$ fF/μ^2, determine the values of R and C for minimum area.

2.7. Assuming the typical processing costs of Table 1.4-2, calculate the cost per minimum size transistor (3 $\mu \times 3\ \mu$), cost per unit for a poly–poly capacitor ($/pF), and cost per unit resistance ($/Ω) for a polysilicon resistor in the CMOS process of Appendix 2B if fabricated on 5 in. wafers.

2.8. A resistor is constructed as shown in Fig. P2.8. If $R_s = 45\ \Omega/\square$, determine the nominal value of this resistor.

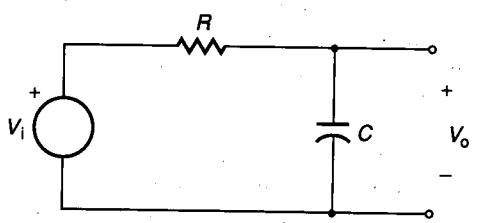


FIGURE P2.6
Problem 2.6.

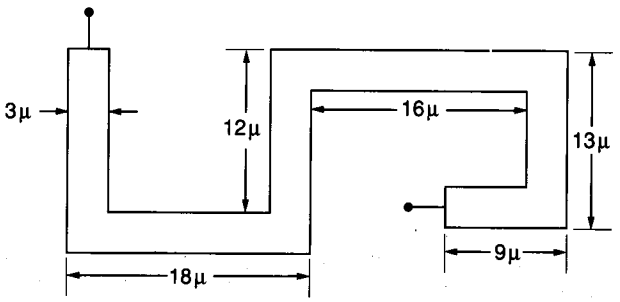


FIGURE P2.8
Problem 2.8.

2.9. In discrete component design the relative "cost" of components is generally as shown in the following list (from lowest to highest). Make a comparable list for the four most common devices in a CMOS process.

Discrete	CMOS
1. Resistor	1.
2. Capacitor	2.
3. Transistor	3.
	4.

2.10. Etching is used in one process step to pattern polysilicon and in a later process step to make contact openings in SiO_2 to gain top contact to residual polysilicon. Why does the contact opening etch not consume the underlying polysilicon?

2.11. Assume that in Fig. P2.11 $W = 5\ \mu, R_\square = 30\ \Omega/\square$ for both poly layers and that the poly–poly capacitance density is 0.7 fF/μ^2.
(*a*) Determine the resistance between nodes A and B.
(*b*) Determine the capacitance between nodes A and C (node B floating).

2.12. Assume the probability of failure for making contacts over a gate of a transistor or on POLY I on top of POLY II is 0.01% because of pinhole problems. If only defects due to gate pinholes are considered, compare the yield of a 50,000-transistor digital circuit with contacts over the gate to a 20-capacitor analog circuit with contacts over the dielectric oxide.

Section 2.3

2.13. Using the design rules of Appendix 2A with $\lambda = 1.5\ \mu$, lay out the circuit shown in Fig. P2.13. Your goal is to minimize effective area minimizing the cost function

$$f = A_1 + \frac{A_2}{10}$$

where A_1 is the area of an imaginary rectangle that encloses all the components

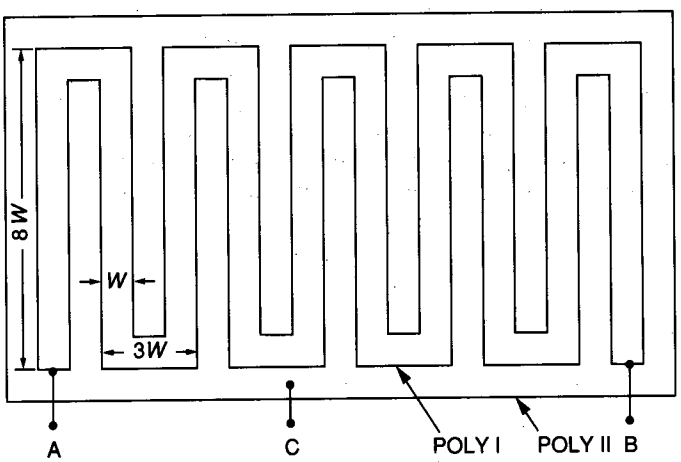


FIGURE P2.11
Problem 2.11.

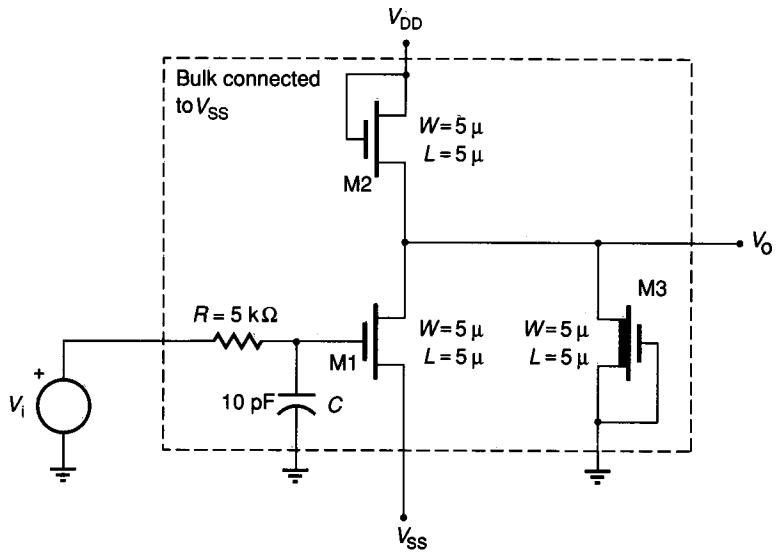


FIGURE P2.13
Problem 2.13.

inside the dashed box and A_2 is the area of a larger rectangle, which encloses the entire circuit (including bonding pads).

2.14. Repeat Problem 2.13 using the design rules of Appendix 2B with $\lambda = 1.5\ \mu$ in the layout of the circuit shown in Fig. P2.14. The substrate is to be connected to V_{DD} and the p-well to ground.

2.15. Repeat Problem 2.13 using the design rules of Appendix 2C with $\lambda = 2.5\ \mu$ in the layout of the circuit shown in Fig. P2.15.

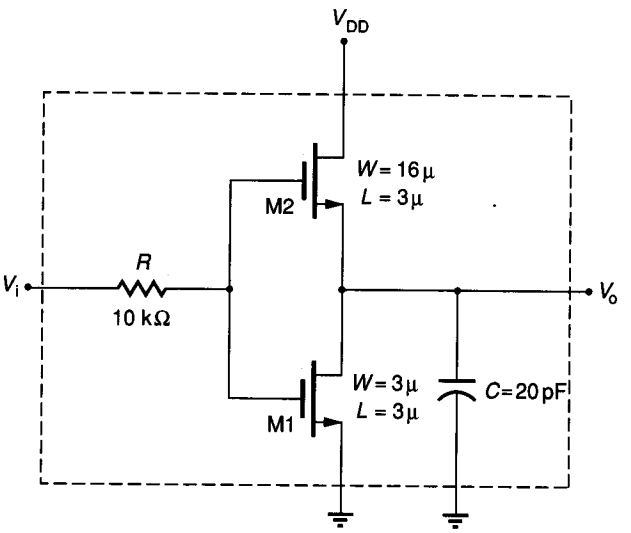


FIGURE P2.14
Problem 2.14.

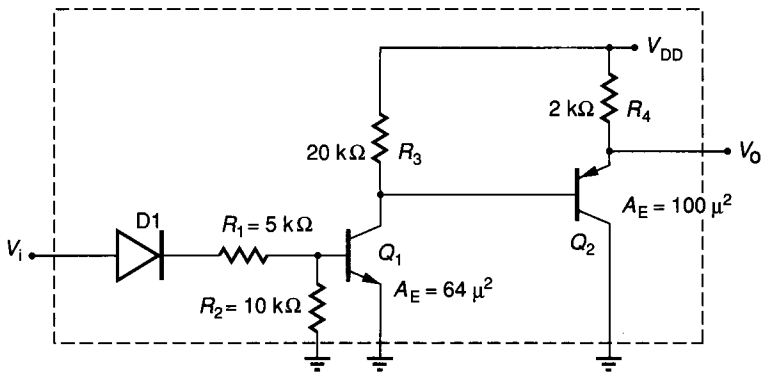


FIGURE P2.15
Problem 2.15.

Section 2.4

2.16. A first-order RC filter with a POLY I resistor and a poly–poly capacitor is shown in Fig. P2.16. What can be said about the following?

(*a*) The variation (3σ) of the 3 db cutoff frequency due to process parameter variations.

(*b*) The variation of the 3 db cutoff frequency due to temperature variations (assume TCR = 0.1%/°C and TCC = 200ppm/°C) over the temperature range 0°C to 70°C?

(*c*) The silicon area required for the circuit.

(*d*) The overall practicality of such a circuit.

Assume the circuit is to be fabricated in the CMOS process characterized by the design rules and process parameters of Appendix B. Be quantitative with your response.

2.17. Modify the layout of the circuit of Fig. 2.2-5, shown in Fig. 2A.1 of Appendix 2A, to use the buried contact instead of metal for connecting the gate of M2 to its source.

2.18. Design and lay out an input pad protection device using the two-diode protection circuit discussed in Sec. 2.2.1C. Use the design rules and process parameters of Appendix 2B.

2.19. The layout of a single-diode input protection scheme is shown in Fig. P2.19*a*.

(*a*) If this is modeled by the circuit in Fig. P2.19*b*, determine the approximate values of R and C. Assume the devices are fabricated in the 3 μ CMOS process characterized by the parameters in Table 2B.4.

(*b*) How does the size of the capacitor compare to that inherent in a normal bonding pad?

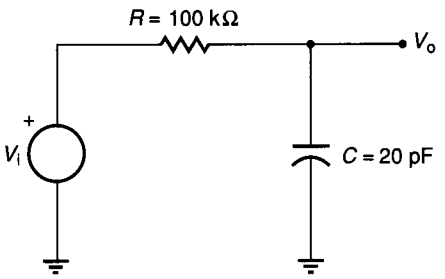


FIGURE P2.16
Problem 2.16.

(*c*) How does the size of the capacitor compare to the typical parasitic capacitance seen on a pin in a dual inline package which is approximately 1.5 pF?

2.20. A layout of a Schmitt trigger circuit that is to be fabricated in the CMOS process of Appendix 2B is shown in Plate 8 in the color plates. Obtain the circuit schematic, including device sizing.

2.21. Repeat Problem 2.20 for the latch circuit of Plate 9.

2.22. Repeat Problem 2.20 for the common source amplifier of Plate 10.

2.23. Repeat Problem 2.20 for the flip-flop of Plate 11.

2.24. Repeat Problem 2.20 for the differential amplifier of Plate 12.

2.25. A die photograph of a CMOS logic circuit fabricated in the CMOS process of Appendix 2B is shown in Plate 13. Obtain the circuit schematic, including device sizing. Assume the minimum metal width in the plate is 5 μ.

2.26. Repeat Problem 2.25 for the logic circuit of Plate 14.

2.27. Repeat Problem 2.25 for the current source circuit of Plate 15.

2.28. Lay out the following circuits using the design rules of Appendix 2B. Try to use good layout techniques and keep the die area to a minimum.
(*a*) Common source amplifier circuit of Fig. P2.28*a*.

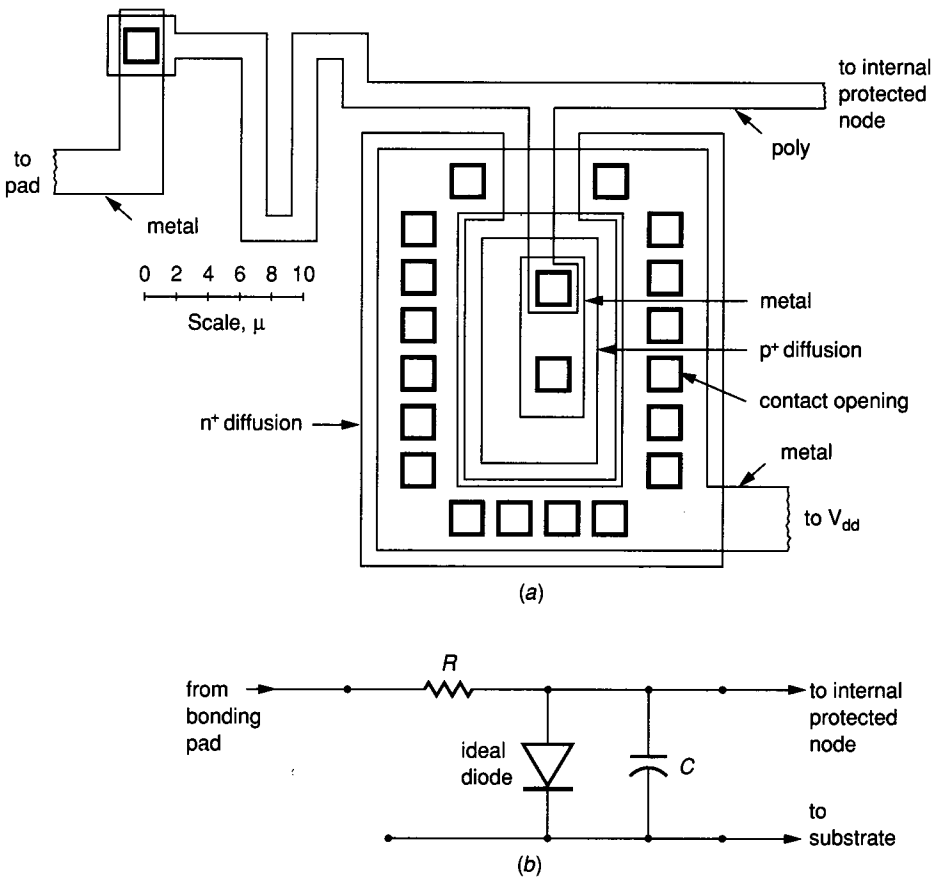
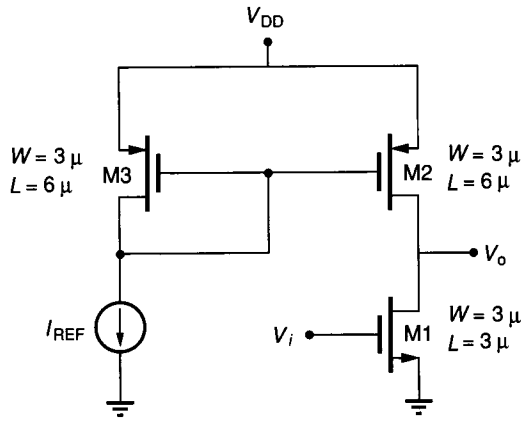
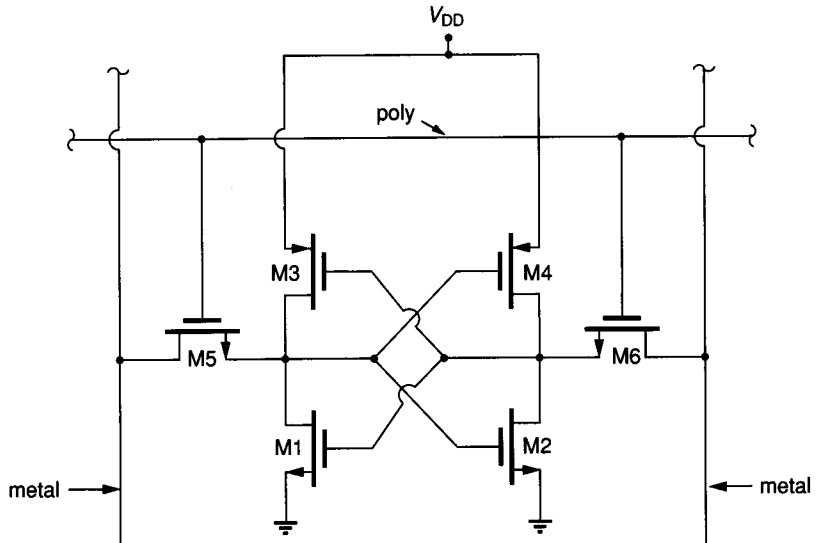


FIGURE P2.19
Single-diode input protection scheme: (*a*) Layout, (*b*) Circuit.

(*b*) Static RAM Cell of Fig. P2.28*b*.
(*c*) Current mirror of Fig. 5.4-12 on p. 349 (device sizes given on p. 348).
(*d*) Op amp of Fig. 6.5-15*a* on p. 480 with device sizes given in Table 6.5-12 on p. 487. Provide bonding pads for external connection of R_{BIAS} and assume C_c is a 2 Pf poly-poly capacitor (see Table 2.B.4).
(*e*) The CMOS Nand gate of Fig. 7.6-14 on p. 554 if all devices are equally sized with $W = L = 3\mu$.



(*a*)



	M1, M2	M3, M4	M5, M6
W	8 μ	3 μ	4 μ
L	2 μ	3 μ	2 μ

(*b*)

FIGURE P2.28
Problem 2.28: (*a*) Common source amplifier circuit, (*b*) Static RAM cell.

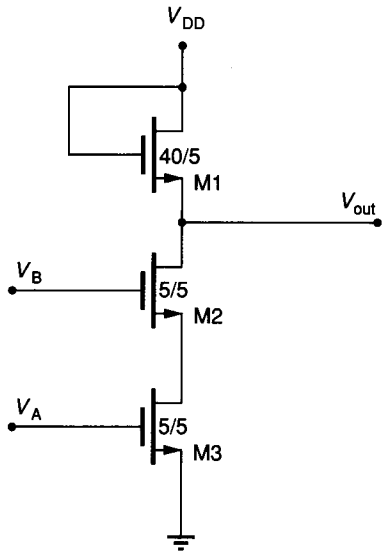


FIGURE P2.29
2-input NAND gate.

2.29. A two-input all enhancement NAND circuit is shown in Fig. P2.29, the layout of which appears in Plate 16. This layout was for the CMOS process of Section 2B of the Appendix. There are one or more layout errors in this layout. Find them and state what needs to be done to correct them. *W/L* values are indicated on the figure.

2.30. Repeat Problem 2.29 for the current source of Fig. P2.30 and the erroneous layout of Plate 17. *W/L* values are indicated on the figure.

2.31. Repeat Problem 2.29 for the differential amplifier of Fig. P2.31 and the erroneous layout of Plate 18. *W/L* values are indicated on the figure.

2.32. Repeat Problem 2.29 for the flip-flop of Fig. P2.32 and the erroneous layout of Plate 19. *W/L* values are indicated on the figure.

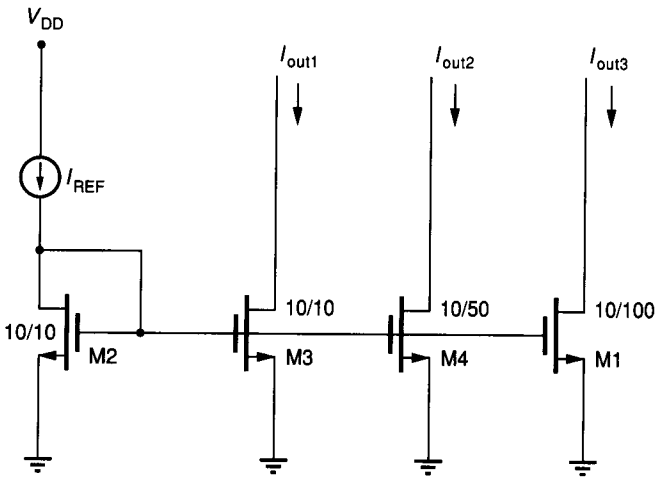


FIGURE P2.30
Current source.

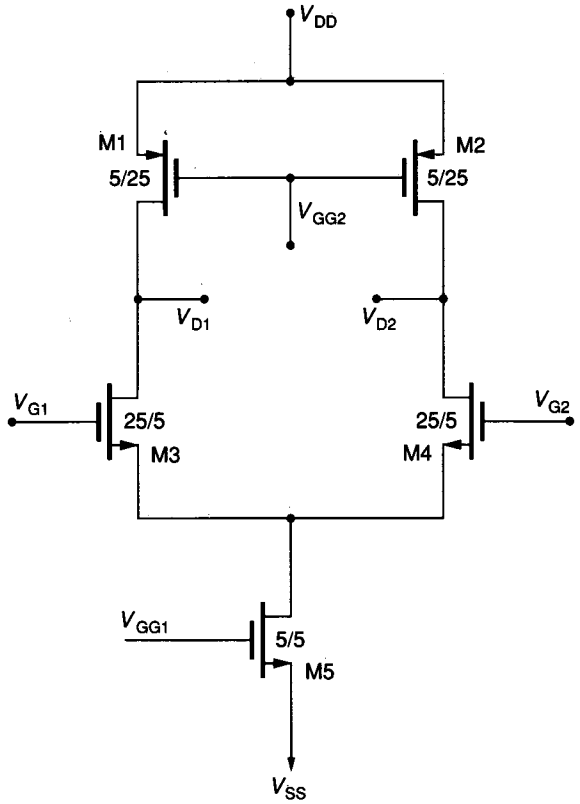


FIGURE P2.31
Differential amplifier.

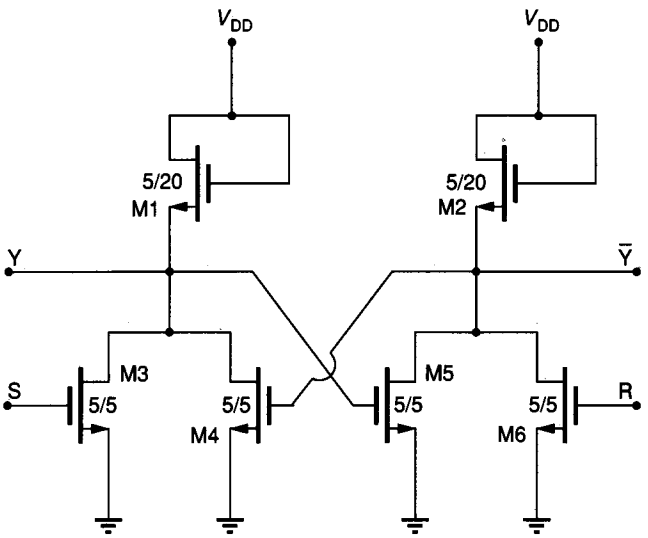


FIGURE P2.32
Flip-flop.

2.33. The NMOS process of Appendix 2A is termed *self-aligned*. Assume that the moat mask was misaligned by 0.75 μ in the positive x direction and perfectly aligned in the y direction relative to the rectangular coordinates shown (in Fig. P2.33). How much misalignment of the gate mask can be tolerated in each of the following directions if the transistors are to remain functional? Assume $\lambda = 1.5\ \mu$ and the features shown in the figure are drawn features.

(*a*) Positive x direction.
(*b*) Positive y direction.
(*c*) Negative x direction.
(*d*) Negative y direction.

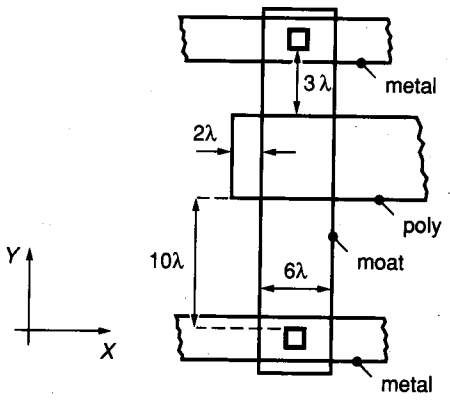


FIGURE P2.33
Mask information for problem 2.33.

APPENDIX 2A
PROCESS CHARACTERIZATION OF A GENERIC NMOS PROCESS

TABLE 2A.1
Process scenario of major process steps in typical NMOS process[a]

1.	Clean wafer	
2.	GROW THIN OXIDE	
3.	Deposit Si_3N_4	
4.	Apply photoresist layer	
5.	PATTERN ANTIMOAT	(MASK #1)
6.	Develop photoresist	
7.	Etch Si_3N_4	
8.	FIELD IMPLANT	
9.	Strip photoresist	
10.	GROW FIELD OXIDE	
11.	Strip Si_3N_4	
12.	Strip thin oxide	
13.	GROW THIN OXIDE	
14.	Channel implant (optional)	
15.	Apply photoresist	
16.	PATTERN DEPLETION TRANSISTOR CHANNELS	(MASK #2)
17.	Develop photoresist	
18.	DEPLETION IMPLANT	
19.	Strip photoresist	
20.	Strip thin oxide	
21.	GROW THIN OXIDE	
	Optional buried contact steps	
	A.1 Apply photoresist	
	A.2 Pattern buried contact	(MASK #A)
	A.3 Develop photoresist	
	A.4 Etch SiO_2 (open via for buried contacts)	
	A.5 Strip photoresist	
22.	POLYSILICON DEPOSITION (POLY I)	
23.	Apply photoresist	
24.	PATTERN POLY I	(MASK #3)
25.	Develop photoresist	
26.	Etch POLY I	
27.	Strip photoresist	
	Optional second polysilicon layer	
	B.1 Strip thin oxide	
	B.2 GROW THIN OXIDE	
	B.3 POLYSILICON DEPOSITION (POLY II)	
	B.4 Apply photoresist	
	B.5 PATTERN POLY II	(MASK #B)
	B.6 Develop photoresist	
	B.7 ETCH POLY II	
	B.8 Strip photoresist	
	B.9 Strip thin oxide	

TABLE 2A.1
(Continued)

28.	CHANNEL DIFFUSION	
29.	Grow oxide	
30.	Apply photoresist	
31.	PATTERN CONTACT OPENINGS	(MASK #4)
32.	Develop photoresist	
33.	ETCH OXIDE	
34.	Strip photoresist	
35.	DEPOSIT METAL	
36.	Apply photoresist	
37.	PATTERN METAL	(MASK #5)
38.	Develop photoresist	
39.	ETCH METAL	
40.	Strip photoresist	
41.	APPLY PASSIVATION	
42.	Apply photoresist	
43.	PATTERN PAD OPENINGS	(MASK #6)
44.	Develop photoresist	
45.	Etch passivation	
46.	Strip photoresist	
47.	ASSEMBLE PACKAGE AND TEST	

[a]Major functional steps shown in capital letters

TABLE 2A.2
Design rules for a typical NMOS process
(See Table 2A.3 in color plates for graphical interpretation)

		Scalable dimension
1.	n^+ Diffusion (moat) (CIF[a] Green, Mask #1)[b]	
	1.1 Width	2λ
	1.2 Spacing	3λ
2.	Implant (CIF Yellow, Mask #2)	
	2.1 Overlap of depletion transistor channel	2λ
	2.2 Spacing to channel of enhancement transistor	1.5λ
3.	Buried contact (CIF Brown, Mask #A)	
	3.1 Spacing to transistor channel	2λ
	3.2 Buried contact overlap of diffusion	2λ
	3.3 Buried overlap of poly or field	λ
	3.4 Buried contact space to unrelated poly or diffusion	2λ

TABLE 2A.2
(Continued)

	Scalable dimension
4. POLY I (CIF Red, Mask #3)	
4.1 Width	2λ
4.2 Spacing	2λ
4.3 Spacing to diffusion	λ
4.4 Overlap of channel diffusion	2λ
4.5 Distance to n^+ diffusion source or drain edge	2λ
5. Contact (CIF Black, Mask #4)	
5.1 Size (exactly)	$2\lambda \times 2\lambda$
5.2 n^+ diffusion overlap of contact	λ
5.3 POLY I overlap of contact	λ
5.4 Contact-to-contact spacing	2λ
5.5 Contact-to-transistor channel spacing	2λ
5.6 Metal overlap of contact	λ
6. Metal (CIF Blue, Mask #5)	
6.1 Width	3λ
6.2 Spacing	3λ
6.3 Maximum current density	1 mA/μ width
6.4 Minimum bonding pad size	100 $\mu \times$ 100 μ
6.5 Minimum probe pad size	75 $\mu \times$ 75 μ
6.6 Minimum pad spacing	50 μ
6.7 Pad to circuitry	40 μ
7. Passivation (CIF Purple, Mask #6)	
7.1 Minimum bonding pad opening	90 $\mu \times$ 90 μ
7.2 Minimum probe pad opening	65 $\mu \times$ 65 μ
8. POLY II (CIF Purple, Mask #B)	
8.1 Width	2λ
8.2 Spacing	2λ
8.3 POLY I overlap of POLY II[c]	λ
8.4 Overlap of contact	2λ

[a]The Caltech Intermediate Format (CIF) is a format for the geometrical database used to describe the layout of an integrated circuit.

[b]Mask numbers are relative to the process scenario of Table 2A.1.

[c]POLY II may not cross a POLY I boundary. POLY I must always be under POLY II.

TABLE 2A.3
Graphical interpretation of NMOS design rules

(See color plate 5 in insert section)

TABLE 2A.4
Process parameters for a typical NMOS process[a]

Parameter	Typical	Tolerance[b]	Units
Square law model parameters			
V_{T0} (threshold voltage)			
Enhancement	0.8	± 0.15	V
Depletion	−3.0	± 0.5	V
K' (conduction factor)			
Enhancement	25	± 20%	$\mu A/V^2$
Depletion	25	± 20%	$\mu A/V^2$
γ (body effect)			
Enhancement	0.4	± 25%	$V^{1/2}$
Depletion	0.4	± 25%	$V^{1/2}$
λ(channel length modulation)	0.01	± 50%	V^{-1}
ϕ (surface potential)	0.6	± 0.1%	V
Process parameters			
μ (channel mobility)	700	±100	$cm^2/(V \cdot s)$
Doping			
Substrate (N_{SUB})	10	±1	$10^{15}/cm^3$
Enhancement n^+	5	±4	$10^{15}/cm^3$
Depletion n^+	1	±0.5	$10^{15}/cm^3$
Channel stop	150	±120	$10^{15}/cm^3$
Physical feature sizes			
T_{OX} (gate oxide thickness)	500	±100	Å
POLY I to POLY II oxide thickness	750	±100	Å
Total lateral diffusion, LD	0.5	±0.25	μ
Width reduction	0.5	±0.25	μ
X_j (n^+ diffusion depth)	0.5	±0.2	μ
Capacitances[c]			
C_{OX} (gate to channel)	0.7	±0.1	fF/μ^2
POLY I to substrate	0.055	±0.01	fF/μ^2
POLY II to substrate	0.06	±0.01	fF/μ^2
Metal to substrate	0.025	±0.005	fF/μ^2
POLY II to POLY I	0.5	±0.1	fF/μ^2
Metal to POLY II	0.05	±0.1	fF/μ^2
Metal to POLY I	0.04	±0.01	fF/μ^2
Metal to moat	0.07	±0.01	fF/μ^2
n^+ diffusion to substrate (bottom)	0.12	±0.03	fF/μ^2
n^+ diffusion to substrate (sidewall)	0.2	±0.1	fF/μ perimeter

TABLE 2A.4
(Continued)

Resistances			
Substrate (resistivity)	25	±20%	Ω · cm
Moat	10	±20%	Ω/□
POLY I (covered by POLY II)	60	±25%	Ω/□
POLY I (uncovered) and POLY II	25	±25%	Ω/□
Metal	0.03	±25%	Ω/□
Contact resistance (per minimum size contact)			
Metal-poly	0.3	±25%	Ω
Metal-n^+ diffusion	1.5	±25%	Ω
Breakdown voltages, leakage currents, migration currents and operating conditions			
Punchthrough voltages (gate oxide and POLY I-POLY II)	>7		V
Diffusion reverse breakdown voltage	>20		V
Metal-field threshold voltage	>10		V
Poly-field threshold voltage	>7		V
Maximum operating voltage	7		V
n^+ diffusion to substrate leakage current	0.1		fA/μ^2
Maximum metal current density	1.0		mA/μ width
Maximum device operating temperature	125		°C
Minimum device operating temperature	−60		°C

[a]Parameter values based upon a 3 μ ($\lambda = 1.5\mu$) process. Most are the same as used for parameter targets for a 3 μ MOSIS process. Some parameters are not self consistent with others listed in the table and the interrelationships established in Chapters 3 and 4 (e.g., $K' \neq \mu C_{OX}$). This is due to the way parameters have been experimentally extracted from test circuits.

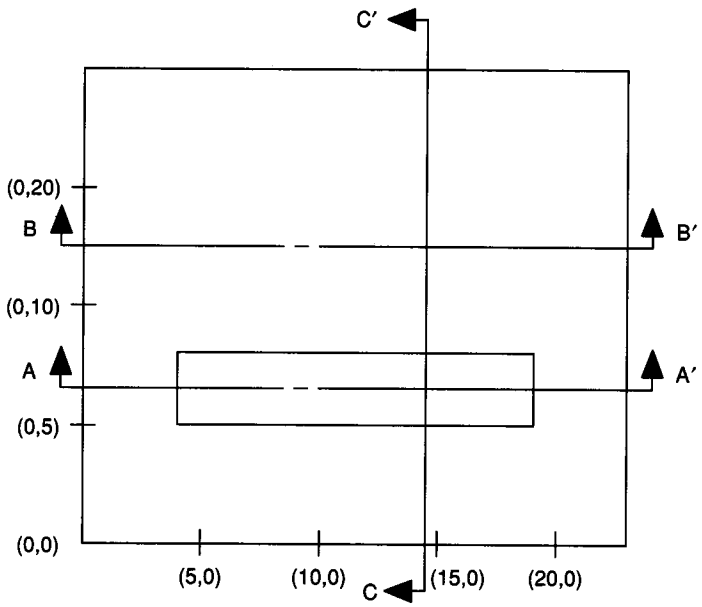
[b]The tolerance is in terms of the absolute value of the parameter relative to processing variations from run to run. Matching characteristics on a die are much better. For example, chip-level matching of V_{TO} is in the 1 to 20 mV range, and K' matching is in the 0.5% to 5% range.[18–22]

[c]Junction capacitances at zero bias.

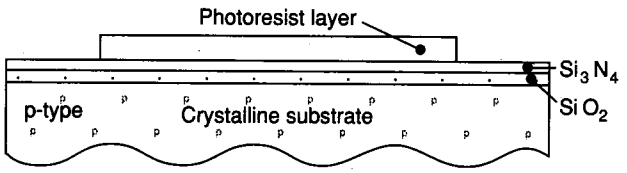
TABLE 2A.5
SPICE MOSFET model parameters of a typical NMOS process (MOSIS)[a]

Parameter (Level 2 model)	Enhancement	Depletion	Units
VTO	1.14	−3.79	V
KP	37.3	32.8	$\mu A/V^2$
GAMMA	0.629	0.372	$V^{1/2}$
PHI	0.6	0.6	V
LAMBDA	3.1E−2	1.0E−6	V^{-1}
CGSO	1.60 E−4	1.60 E−4	fF/μ width
CGDO	1.60 E−4	1.60 E−4	fF/μ width
CGBO	1.70 E−4	1.70 E−4	fF/μ width
RSH	25.4	25.4	Ω/□
CJ	1.1E−4	1.1E−4	pF/μ^2
MJ	0.5	0.5	
CJSW	5.0 E−4	5.0 E−4	pF/μ perimeter
MJSW	0.33	0.33	
TOX	544	544	Å
NSUB	2.09E15	1.0E16	$1/cm^3$
NSS	0	0	$1/cm^2$
NFS	1.90E12	4.31E12	$1/cm^2$
TPG	1	1	
XJ	1.31	0.6	μ
LD	0.826	1.016	μ
UO	300	900	$cm^2/(V \cdot s)$
UCRIT	1.0E6	0.805E6	V/cm
UEXP	1.001E−3	1.001E−3	
VMAX	1.0E5	6.75E5	m/s
NEFF	1.001E−2	1.001E−2	
DELTA	1.16	2.80	

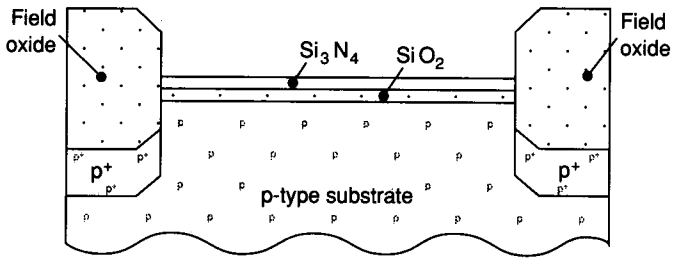
[a] The SPICE parameters were obtained by assuming them to be empirical parameters and then fitting measured device characteristics to the mathematical equations which comprise the model by using a numerical optimization algorithm. This approach gives good fit to the model but causes a deviation from the typical parameters of Table 2A.4 and results in parameter relationships which may not be self-consistent with some of the fundamental relationships developed in Chapters 3 and 4.



Mask #1 Moat Definition



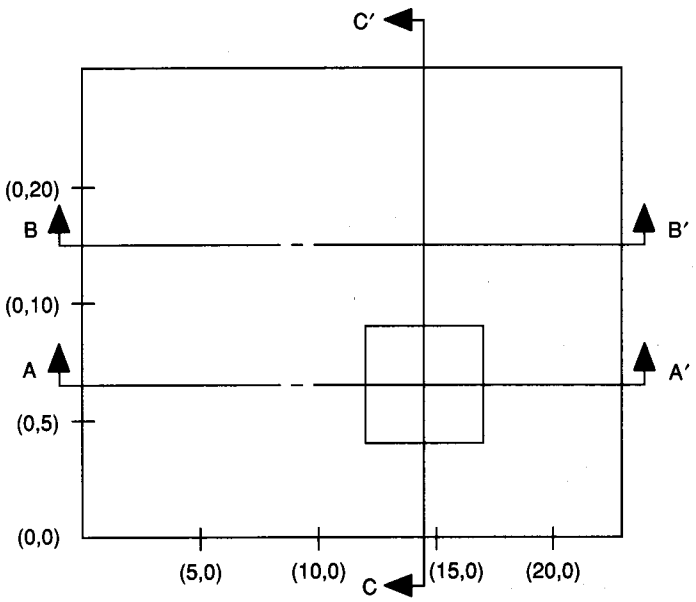
A-A′ Following Moat Patterning



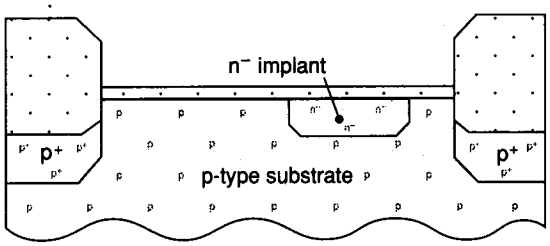
A-A′ Following Field Implant and Oxide Growth

(*a*)

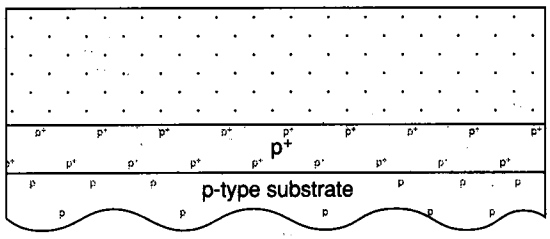
FIGURE 2A.1
Process description for the NMOS process.



Mask #2 Depletion Implant



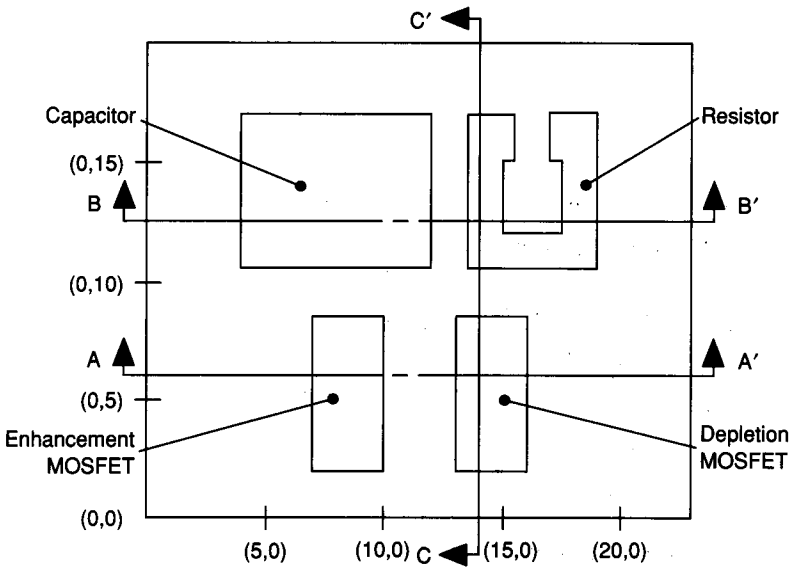
A-A′ Following Depletion Implant



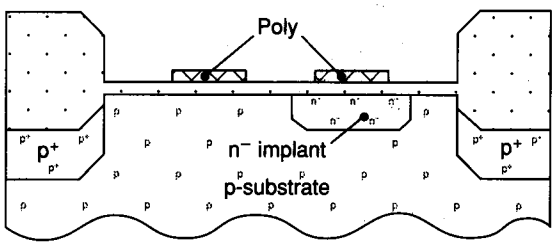
B-B′ Following Depletion Implant

(*b*)

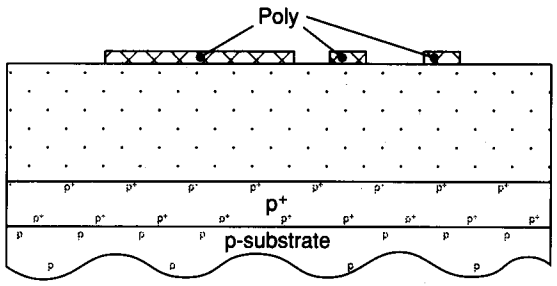
FIGURE 2A.1
(Continued)



Mask #3 POLY 1



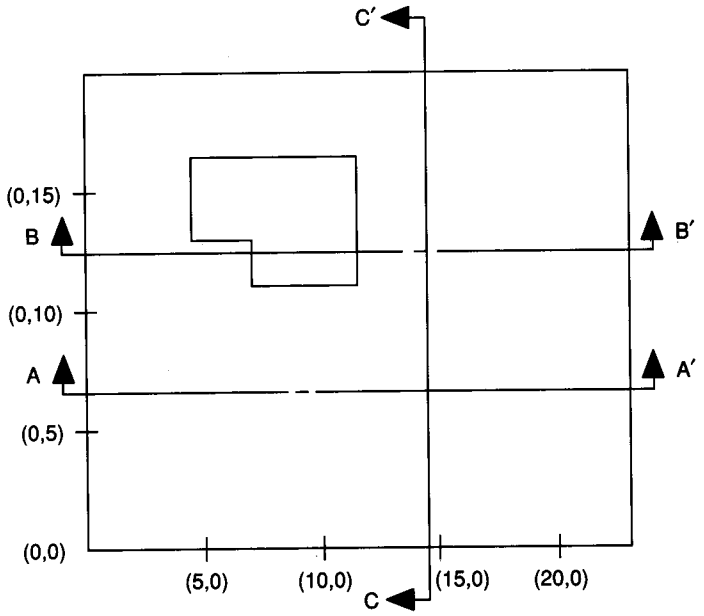
A-A′ Following Patterning POLY I



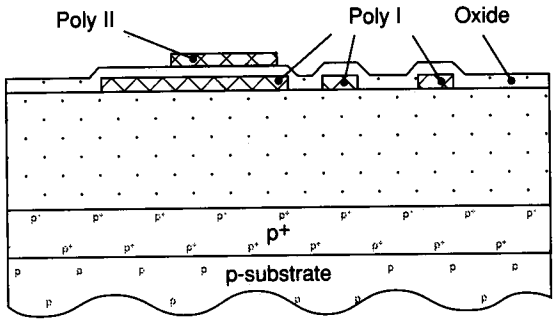
B-B′ Following Patterning POLY I

(*c*)

FIGURE 2A.1
(Continued)



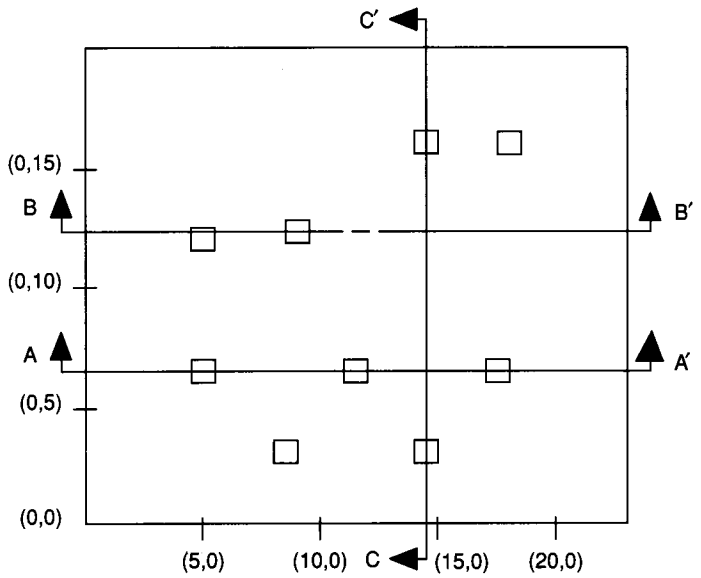
Mask B POLY II



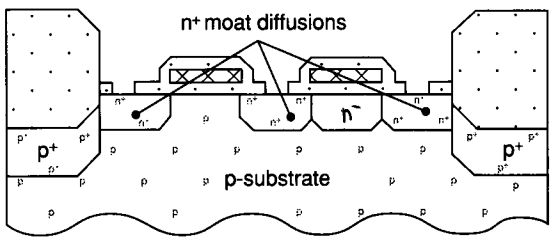
B-B′ Following Patterning POLY II

(*d*)

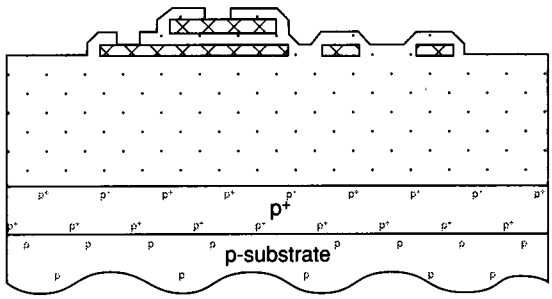
FIGURE 2A.1
(Continued)



Mask #4 Contact Openings



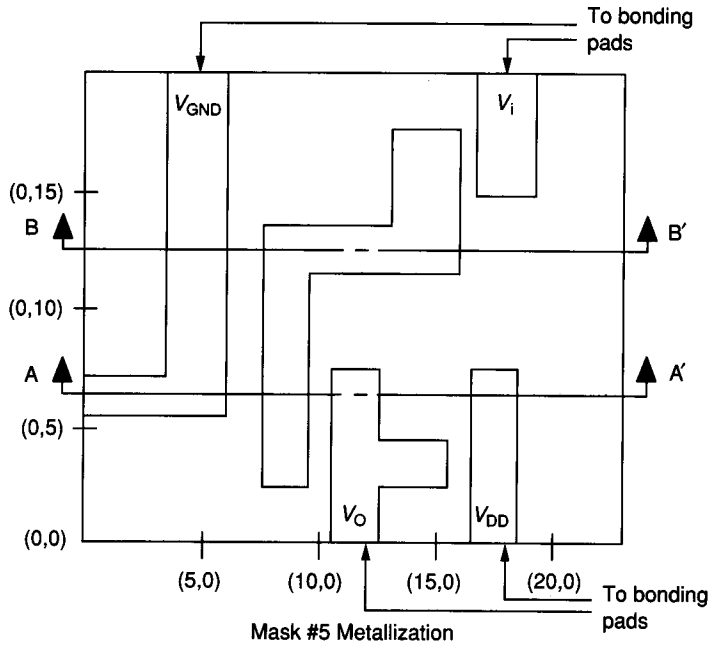
A-A′ Following Moat Diffusion and Contact Openings

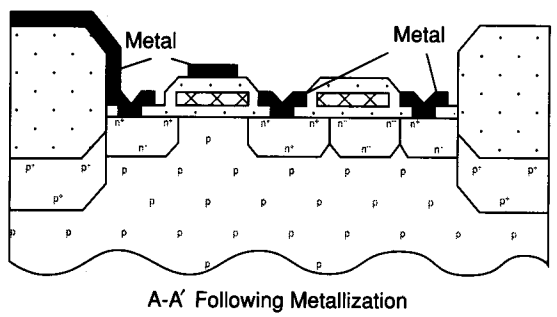


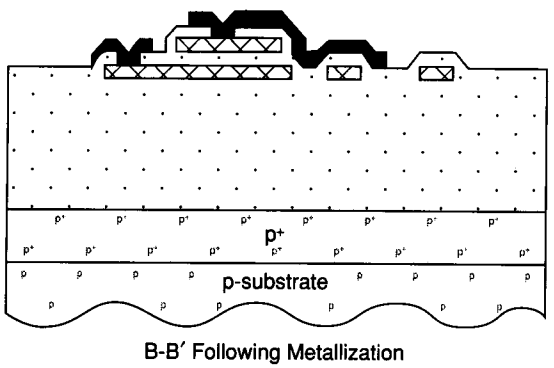
B-B′ Following Moat Diffusion and Contact Openings

(*e*)

FIGURE 2A.1
(Continued)







(*f*)

FIGURE 2A.1
(Continued)

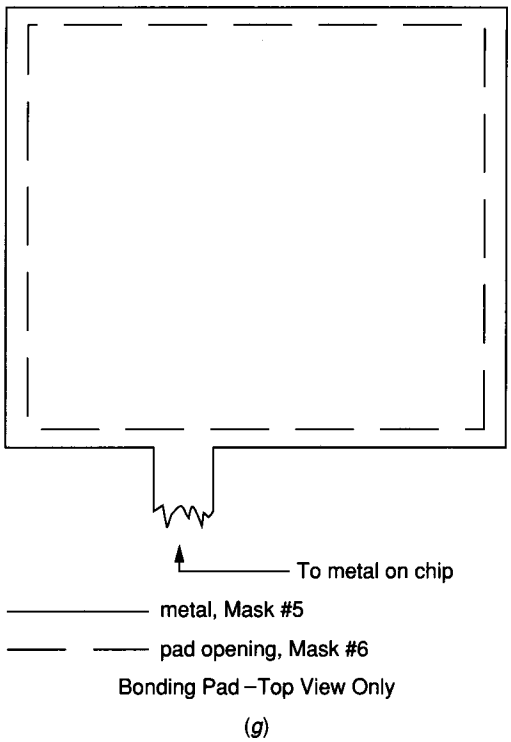


FIGURE 2A.1
(Continued)

APPENDIX 2B
PROCESS CHARACTERIZATION OF A GENERIC CMOS PROCESS

TABLE 2B.1
Process scenario of major process steps in typical p-well CMOS process[a]

1.	Clean wafer	
2.	GROW THIN OXIDE	
3.	Apply photoresist	
4.	PATTERN P-WELL	(MASK #1)
5.	Develop photoresist	
6.	Deposit and diffuse p-type impurities	
7.	Strip photoresist	
8.	Strip thin oxide	
9.	Grow thin oxide	
10.	Apply layer of Si_3N_4	
11.	Apply photoresist	
12.	PATTERN Si_3N_4 (active area definition)	(MASK #2)
13.	Develop photoresist	
14.	Etch Si_3N_4	
15.	Strip photoresist	
	Optional field threshold voltage adjust	
	A.1 Apply photoresist	
	A.2 PATTERN ANTIMOAT IN SUBSTRATE	(MASK #A1)
	A.3 Develop photoresist	
	A.4 FIELD IMPLANT (n-type)	
	A.5 Strip photoresist	
16.	GROW FIELD OXIDE	
17.	Strip Si_3N_4	
18.	Strip thin oxide	
19.	GROW GATE OXIDE	
20.	POLYSILICON DEPOSITION (POLY I)	
21.	Apply photoresist	
22.	PATTERN POLYSILICON	(MASK #3)
23.	Develop photoresist	
24.	ETCH POLYSILICON	
25.	Strip photoresist	
	Optional steps for double polysilicon process	
	B.1 Strip thin oxide	
	B.2 GROW THIN OXIDE	
	B.3 POLYSILICON DEPOSITION (POLY II)	
	B.4 Apply photoresist	
	B.5 PATTERN POLYSILICON	(MASK #B1)
	B.6 Develop photoresist	
	B.7 ETCH POLYSILICON	
	B.8 Strip photoresist	
	B.9 Strip thin oxide	

TABLE 2B.1
(Continued)

26.	Apply photoresist	
27.	PATTERN P-CHANNEL DRAINS AND SOURCES AND P^+ GUARD RINGS (p-well ohmic contacts)	(MASK #4)
28.	Develop photoresist	
29.	p^+ IMPLANT	
30.	Strip photoresist	
31.	Apply photoresist	
32.	PATTERN N-CHANNEL DRAINS AND SOURCES AND N^+ GUARD RINGS (top ohmic contact to substrate)	(MASK #5)
33.	Develop photoresist	
34.	n^+ IMPLANT	
35.	Strip photoresist	
36.	Strip thin oxide	
37.	Grow oxide	
38.	Apply photoresist	
39.	PATTERN CONTACT OPENINGS	(MASK #6)
40.	Develop photoresist	
41.	Etch oxide	
42.	Strip photoresist	
43.	APPLY METAL	
44.	Apply photoresist	
45.	PATTERN METAL	(MASK #7)
46.	Develop photoresist	
47.	Etch metal	
48.	Strip photoresist	
	Optional steps for double metal process	
	C.1 Strip thin oxide	
	C.2 DEPOSIT INTERMETAL OXIDE	
	C.3 Apply photoresist	
	C.4 PATTERN VIAS	(MASK #C1)
	C.5 Develop photoresist	
	C.6 Etch oxide	
	C.7 Strip photoresist	
	C.8 APPLY METAL (Metal 2)	
	C.9 Apply photoresist	
	C.10 PATTERN METAL	(MASK #C2)
	C.11 Develop photoresist	
	C.12 Etch metal	
	C.13 Strip photoresist	
49.	APPLY PASSIVATION	
50.	Apply photoresist	
51.	PATTERN PAD OPENINGS	(MASK #8)
52.	Develop photoresist	
53.	Etch passivation	
54.	Strip photoresist	
55.	ASSEMBLE, PACKAGE AND TEST	

[a]Major functional steps shown in capital letters.

TABLE 2B.2
Design rules for a typical p-well CMOS process (See Table 2B.3 in color plates for graphical interpretation)

		Dimensions	
		Microns	**Scalable**
1.	p-well (CIF Brown, Mask #1[a])		
1.1	Width	5	4λ
1.2	Spacing (different potential)	15	10λ
1.3	Spacing (same potential)	9	6λ
2.	Active (CIF Green, Mask #2)		
2.1	Width	4	2λ
2.2	Spacing	4	2λ
2.3	p^+ active in n-subs to p-well edge	8	6λ
2.4	n^+ active in n-subs to p-well edge	7	5λ
2.5	n^+ active in p-well to p-well edge	4	2λ
2.6	p^+ active in p-well to p-well edge	1	λ
3.	Poly (POLY I) (CIF Red, Mask #3)		
3.1	Width	3	2λ
3.2	Spacing	3	2λ
3.3	Field poly to active	2	λ
3.4	Poly overlap of active	3	2λ
3.5	Active overlap of poly	4	2λ
4.	p^+ select (CIF Orange, Mask #4)		
4.1	Overlap of active	2	λ
4.2	Space to n^+ active	2	λ
4.3	Overlap of channel[b]	3.5	2λ
4.4	Space to channel[b]	3.5	2λ
4.5	Space to p^+ select	3	2λ
4.6	Width	3	2λ
5.	Contact[c] (CIF Purple, Mask #6)		
5.1	Square contact, exactly	3 × 3	2λ × 2λ
5.2	Rectangular contact, exactly	3 × 8	2λ × 6λ
5.3	Space to different contact	3	2λ
5.4	Poly overlap of contact	2	λ
5.5	Poly overlap in direction of metal 1	2.5	2λ
5.6	Space to channel	3	2λ
5.7	Metal 1 overlap of contact	2	λ
5.8	Active overlap of contact	2	λ
5.9	p^+ select overlap of contact	3	2λ
5.10	Subs./well shorting contact, exactly	3 × 8	2λ × 6λ
6.	Metal 1[d] (CIF Blue, Mask #7)		
6.1	Width	3	2λ
6.2	Spacing	4	3λ
6.3	Maximum current density	0.8 mA/μ	0.8 mA/μ

TABLE 2B.2
(Continued)

		Dimensions	
		Microns	**Scalable**
7.	Via [e](CIF Purple Hatched, Mask #C1)		
7.1	Size, exactly	3 × 3	2λ × 2λ
7.2	Separation	3	2λ
7.3	Space to poly edge	4	2λ
7.4	Space to contact	3	2λ
7.5	Overlap by metal 1	2	λ
7.6	Overlap by metal 2	2	λ
7.7	Space to active edge	3	2λ
8.	Metal 2 (CIF Orange Hatched, Mask #C2)		
8.1	Width	5	3λ
8.2	Spacing	5	3λ
8.3	Bonding pad size	100 × 100	100 μ × 100 μ
8.4	Probe pad size	75 × 75	75 μ × 75 μ
8.5	Bonding pad separation	50	50 μ
8.6	Bonding to probe pad	30	30 μ
8.7	Probe pad separation	30	30 μ
8.8	Pad to circuitry	40	40 μ
8.9	Maximum current density	0.8 mA/μ	0.8 mA/μ
9.	Passivation[f] (CIF Purple Dashed, Mask #8)		
9.1	Bonding pad opening	90 × 90	90 μ × 90 μ
9.2	Probe pad opening	65 × 65	65 μ × 65 μ
10.	Metal 2 crossing coincident metal 1 and poly[g]		
10.1	Metal 1 to poly edge spacing when crossing metal 2	2	λ
10.2	Rule domain	2	λ
11.	Electrode (POLY II)[h] (CIF Purple Hatched, Mask #A1)		
11.1	Width	3	2λ
11.2	Spacing	3	2λ
11.3	POLY I overlap of POLY II	2	λ
11.4	Space to contact	3	2λ

[a]Mask numbers are relative to the process scenario of Table 2B.1. CIF format discussed in footnote of Table 2A.2.
[b]Add 2.5 microns for a source/drain width of 3μ for worst-case mask misalignment.
[c]No contact to poly inside active.
[d]For single metal process, pads are made with metal 1 following design rules 8.3–8.8.
[e]Via must be on a flat surface; metal 1 must be under a via.
[f]There must be metal 2 under the pad openings in a double-metal process.
[g]Objective: Avoidance of too large a step for metal 2.
[h]POLY I must always be under POLY II.

TABLE 2B.3
Graphical interpretation of CMOS design rules.

(See color plate 6 in insert section)

TABLE 2B.4
Process parameters for a typical[a] p-well CMOS process

	Typical	Tolerance[b]	Units
Square law model parameters			
V_{T0} (threshold voltage)			
n-channel (V_{TN0})	0.75	± 0.25	V
p-channel (V_{TP0})	−0.75	± 0.25	V
K'(conduction factor)			
n-channel	24	± 6	$\mu A/V^2$
p-channel	8	± 1.5	$\mu A/V^2$
γ(body effect)			
n-channel	0.8	± 0.4	$V^{1/2}$
p-channel	0.4	± 0.2	$V^{1/2}$
λ(channel length modulation)			
n-channel	0.01	± 50%	V^{-1}
p-channel	0.02	± 50%	V^{-1}
ϕ(surface potential)			
n- and p-channel	0.6	± 0.1	V
Process parameters			
μ (channel mobility)			
n-channel	710		$cm^2/(V \cdot s)$
p-channel	230		$cm^2/(V \cdot s)$
Doping[c]			
n^+ active	5	±4	$10^{18}/cm^3$
p^+ active	5	±4	$10^{17}/cm^3$
p-well	5	±2	$10^{16}/cm^3$
n-substrate	1	±0.1	$10^{16}/cm^3$
Physical feature sizes			
T_{OX} (gate oxide thickness)	500	± 100	Å
Total lateral diffusion			
n-channel	0.45	± 0.15	μ
p-channel	0.6	± 0.3	μ
Diffusion depth			
n^+ diffusion	0.45	± 0.15	μ
p^+ diffusion	0.6	± 0.3	μ
p-well	3.0	± 30%	μ
Insulating layer separation			
POLY I to POLY II	800	± 100	Å
Metal 1 to Substrate	1.55	± 0.15	μ
Metal 1 to Diffusion	0.925	± 0.25	μ
POLY I to Substrate (POLY I on field oxide)	0.75	± 0.1	μ
Metal 1 to POLY I	0.87	± 0.7	μ
Metal 2 to Substrate	2.7	± 0.25	μ
Metal 2 to Metal I	1.2	± 0.1	μ
Metal 2 to POLY I	2.0	± 0.07	μ

Note: $K' = \mu C_{ox}$ $24E-6 \neq (710)(69) \simeq 49.7E-6$

TABLE 2B.4
(Continued)

	Typical	Tolerance[b]	Units
Capacitances[d]			
C_{OX} (gate oxide capacitance, n- and p-channel)	0.7	±0.1	fF/μ^2
POLY I to substrate, poly in field	0.045	±0.01	fF/μ^2
POLY II to substrate, poly in field	0.045	±0.01	fF/μ^2
Metal 1 to substrate, metal in field	0.025	±0.005	fF/μ^2
Metal 2 to substrate, metal in field	0.014	±0.002	fF/μ^2
POLY I to POLY II	0.44	±0.05	fF/μ^2
POLY I to Metal 1	0.04	±0.01	fF/μ^2
POLY I to Metal 2	0.039	±0.003	fF/μ^2
Metal 1 to Metal 2	0.035	±0.01	fF/μ^2
Metal 1 to diffusion	0.04	±0.01	fF/μ^2
Metal 2 to diffusion	0.02	±0.005	fF/μ^2
n^+ diffusion to p-well (junction, bottom)	0.33	±0.17	fF/μ^2
n^+ diffusion sidewall (junction, sidewall)	2.6	±0.6	fF/μ
p^+ diffusion to substrate (junction, bottom)	0.38	±0.12	fF/μ^2
p^+ diffusion sidewall (junction, sidewall)	3.5	±2.0	fF/μ
p-well to substrate (junction, bottom)	0.2	±0.1	fF/μ^2
p-well sidewall (junction, sidewall)	1.6	±1.0	fF/μ
Resistances			
Substrate	25	±20%	Ω-cm
p-well	5000	±2500	Ω/□
n^+ diffusion	35	±25	Ω/□
p^+ diffusion	80	±55	Ω/□
Metal	0.003	±25%	Ω/□
Poly	25	±25%	Ω/□
Metal 1–Metal 2 via (3 μ × 3 μ contact)	<0.1		Ω
Metal 1 contact to POLY I (3 μ × 3 μ contact)	<10		Ω
Metal 1 contact to n^+ or p^+ diffusion (3 μ × 3 μ contact)	<5		Ω
Breakdown voltages, leakage currents, migration currents and operating conditions			
Punchthrough voltages (Gate oxide, POLY I to POLY II)	>10		V
Diffusion reverse breakdown voltage	>10		V
p-well to substrate reverse breakdown voltage	>20		V
Metal 1 in field threshold voltage	>10		V
Metal 2 in field threshold voltage	>10		V
Poly-field threshold voltage	>10		V
Maximum operating voltage	7.0		V
n^+ diffusion to p-well leakage current	0.25		fA/μ^2
p^+ diffusion to substrate leakage current	0.25		fA/μ^2
p-well leakage current	0.25		fA/μ^2
Maximum metal current density	0.8		mA/μ width
Maximum device operating temperature	200		°C

[a] Parameters based upon a 3μ ($\lambda = 1.5\mu$) CMOS process.

[b] The tolerance is in terms of the absolute value of the parameter relative to processing variations from run to run. Matching characteristics on a die are much better. For example, chip-level matching of V_{T0} is in the 1 mV to 20 mV range, and K′ matching is in the 0.5% to 5% range.[18–22]

[c] Impurity concentration varies with depth.

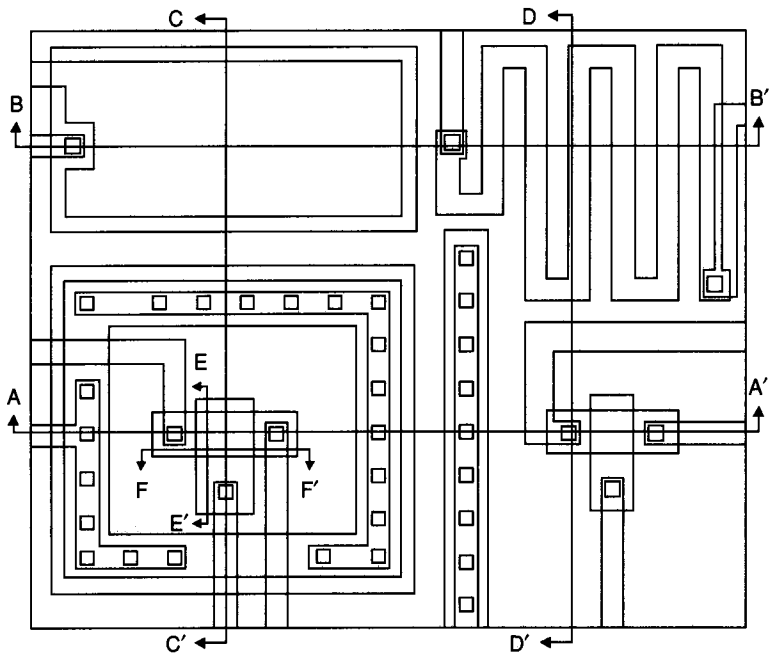
[d] Junction capacitances at zero bias.

TABLE 2B.5
SPICE MOSFET model parameters of a typical p-well CMOS process (MOSIS[a])

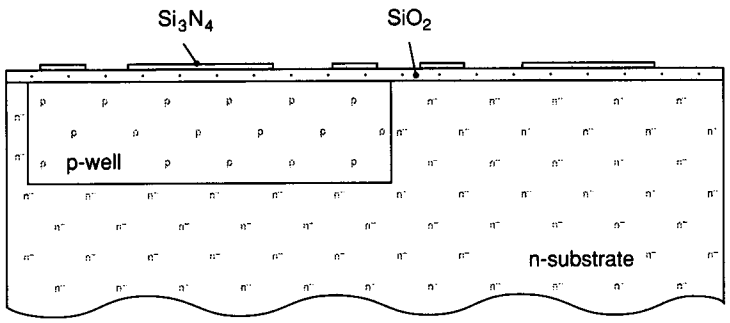
Parameter (Level 2 model)	n-channel	p-channel	Units
VTO	0.827	−0.895	V
KP	32.87	15.26	$\mu A/V^2$
GAMMA	1.36	0.879	$V^{1/2}$
PHI	0.6	0.6	V
LAMBDA	1.605E−2	4.709E−2	V^{-1}
CGSO	5.2E−4	4.0E−4	fF/μ width
CGDO	5.2E−4	4.0E−4	fF/μ width
RSH	25	95	$\Omega/\square$
CJ	3.2E−4	2.0E−4	p/fF/μ^2
MJ	0.5	0.5	
CJSW	9.0E−4	4.5E−4	p/fF/μ perimeter
MJSW	0.33	0.33	
TOX	500	500	Å
NSUB	1.0E16	1.12E14	$1/cm^3$
NSS	0	0	$1/cm^2$
NFS	1.235E12	8.79E11	$1/cm^2$
TPG	1	−1	
XJ	0.4	0.4	μ
LD	0.28	0.28	μ
UO	200	100	$cm^2/(V \cdot s)$
UCRIT	9.99E5	1.64E4	V/cm
UEXP	1.001E−3	0.1534	
VMAX	1.0E5	1.0E5	m/s
NEFF	1.001E−2	1.001E−2	
DELTA	1.2405	1.938	

verify units before final corrections

[a] The SPICE parameters were obtained by assuming them to be empirical parameters and then fitting measured device characteristics to the mathematical equations which comprise the model by using a numerical optimization algorithm. This approach gives good fit to the model but causes a deviation from the typical parameters of Table 2A.4 and results in parameter relationships which may not be self-consistent with some of the fundamental relationships developed in Chapters 3 and 4.



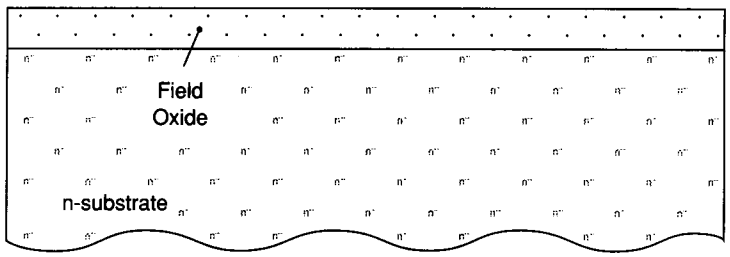
(*a*) Top View (color version appears in plate 3)



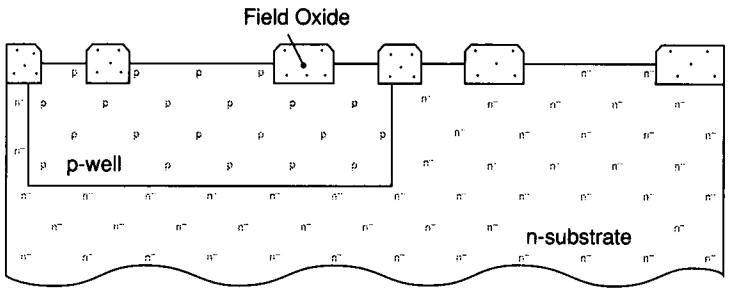
A-A′ Following Moat Patterning

(*b*)

FIGURE 2B.1
Process description for the CMOS process.

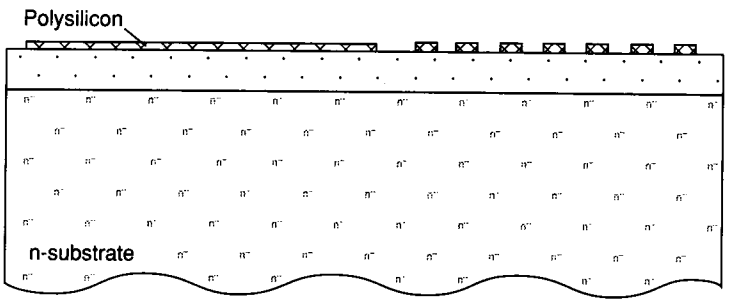


B-B Following Field Oxide Growth

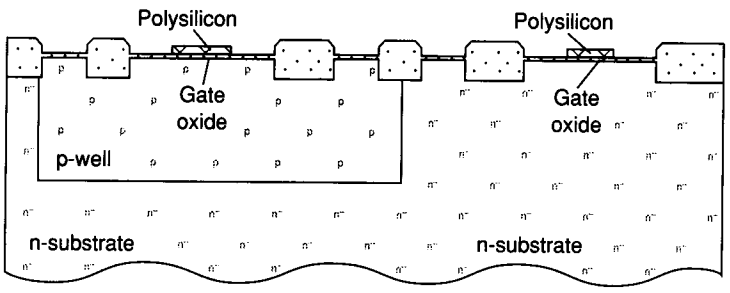


A-A′ Following Field Oxide Growth

(*c*)



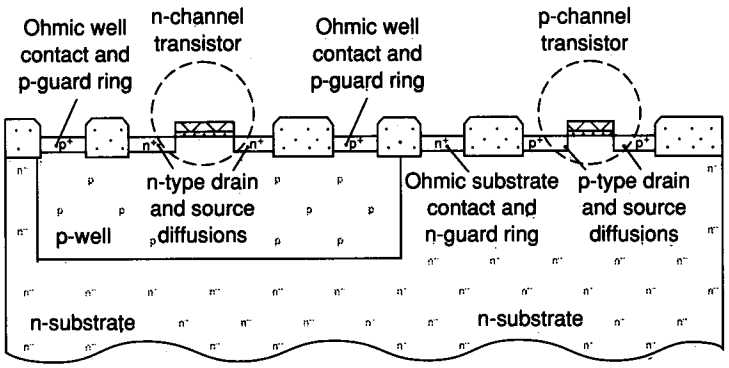
B-B′ Following Polysilicon Patterning



A-A′ Following Polysilicon Patterning

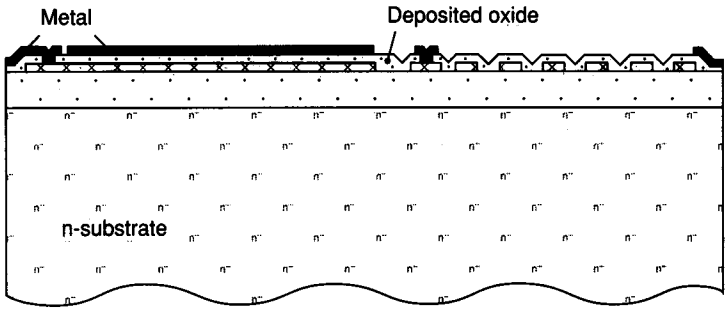
(*d*)

FIGURE 2B.1
(Continued)

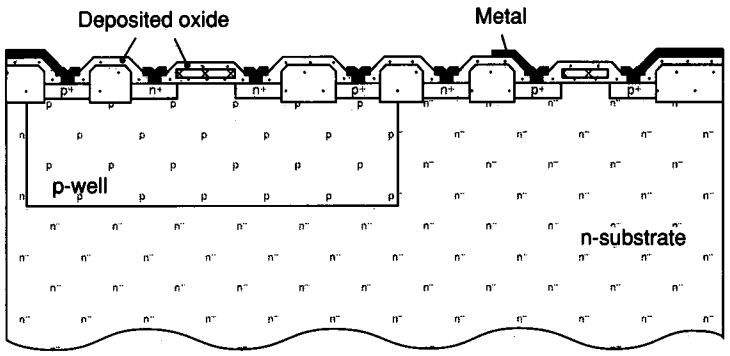


A-A′ Following p+ and n+ Diffusions

(*e*)



B-B′ Following Metal Patterning



A-A′ Following Metal Patterning

(*f*)

FIGURE 2B.1
(Continued)

APPENDIX 2C
PROCESS CHARACTERIZATION OF A GENERIC BIPOLAR PROCESS

TABLE 2C.1
Process scenario of major process steps in typical bipolar process[a]

1.	Clean wafer (p-type)	
2.	GROW THIN OXIDE	
3.	Apply photoresist	
4.	PATTERN n^+ BURIED LAYER	(MASK #1)
5.	Develop photoresist	
6.	DEPOSITION AND DIFFUSION OF n-BURIED LAYER	
7.	Strip photoresist	
8.	Strip oxide	
9.	GROW EPITAXIAL LAYER (n-type)	
10.	Grow oxide	
11.	Apply photoresist	
12.	PATTERN p^+ ISOLATION REGIONS	(MASK #2)
13.	Develop photoresist	
14.	Etch oxide	
15.	DEPOSITION AND DIFFUSION OF p^+ ISOLATION	
16.	Strip photoresist	
17.	Grow oxide	
	Optional high-resistance p-diffusion	
	A.1 Apply photoresist	
	A.2 PATTERN p-RESISTORS	(MASK #A)
	A.3 Develop photoresist	
	A.4 Etch oxide	
	A.5 DEPOSITION AND DIFFUSION OF p-RESISTORS	
	A.6 Strip photoresist	
	A.7 Grow oxide	
18.	Apply photoresist	
19.	PATTERN BASE REGIONS	(MASK #3)
20.	Develop photoresist	
21.	Etch oxide	
22.	DEPOSITION AND DIFFUSION OF p-TYPE BASE	
23.	Strip photoresist	
24.	Grow oxide	
25.	Apply photoresist	
26.	PATTERN n-TYPE EMITTER REGIONS	(MASK #4)
27.	Develop photoresist	
28.	Etch Oxide	
29.	n^+ DEPOSITION AND DIFFUSION	
30.	Strip photoresist	
31.	Grow oxide	
32.	Apply photoresist	
33.	PATTERN CONTACT OPENINGS	(MASK #5)
34.	Develop photoresist	
35.	Etch oxide	
36.	Strip Photoresist	
37.	APPLY METAL	
38.	Apply photoresist	
39.	PATTERN METAL	(MASK #6)

TABLE 2C.1
(Continued)

40.	Develop photoresist	
41.	ETCH METAL	
42.	Strip photoresist	
43.	APPLY PASSIVATION	
44.	Apply photoresist	
45.	PATTERN PAD OPENINGS	(MASK #7)
46.	Develop photoresist	
47.	Etch passivation	
48.	Strip photoresist	
49.	ASSEMBLE, PACKAGE, AND TEST	

[a]Major functional steps shown in capital letters.

TABLE 2C.2
Design rules for a typical bipolar process ($\lambda = 2.5\ \mu$)
(See Table 2C.3 in color plates for graphical interpretation)

	Dimension
1. n^+ buried collector diffusion (Yellow, Mask #1)	
1.1 Width	3λ
1.2 Overlap of p-base diffusion (for vertical npn)	2λ
1.3 Overlap of n^+ emitter diffusion (for collector contact of vertical npn)	2λ
1.4 Overlap of p-base diffusion (for collector and emitter of lateral pnp)	2λ
1.5 Overlap of n^+ emitter diffusion (for base contact of lateral pnp)	2λ
2. Isolation diffusion (Orange, Mask #2)	
2.1 Width	4λ
2.2 Spacing	24λ
2.3 Distance to n^+ buried collector	14λ
3. p-base diffusion (Brown, Mask #3)	
3.1 Width	3λ
3.2 Spacing	5λ
3.3 Distance to isolation diffusion	14λ
3.4 Width (resistor)	3λ
3.5 Spacing (as resistor)	3λ
4. n^+ emitter diffusion (Green, Mask #4)	
4.1 Width	3λ
4.2 Spacing	3λ
4.3 p-base diffusion overlap of n^+ emitter diffusion (emitter in base)	2λ
4.4 Spacing to isolation diffusion (for collector contact)	12λ
4.5 Spacing to p-base diffusion (for base contact of lateral pnp)	6λ
4.6 Spacing to p-base diffusion (for collector contact of vertical npn)	6λ
5. Contact (Black, Mask #5)	
5.1 Size (exactly)	$4\lambda \times 4\lambda$
5.2 Spacing	2λ
5.3 Metal overlap of contact	λ
5.4 n^+ emitter diffusion overlap of contact	2λ
5.5 p-base diffusion overlap of contact	2λ
5.6 p-base to n^+ emitter	3λ
5.7 Spacing to isolation diffusion	4λ

TABLE 2C.2
(Continued)

	Dimension
6. Metalization (Blue, Mask #6)	
6.1 Width	2λ
6.2 Spacing	2λ
6.3 Bonding pad size	100 μ × 100 μ
6.4 Probe pad size	75 μ × 75 μ
6.5 Bonding pad separation	50 μ
6.6 Bonding to probe pad	30 μ
6.7 Probe pad separation	30 μ
6.8 Pad to circuitry	40 μ
6.9 Maximum current density	0.8 mA/μ width
7. Passivation (Purple, Mask #7)	
7.1 Minimum bonding pad opening	90 μ × 90 μ
7.2 Minimum probe pad opening	65 μ × 65 μ

TABLE 2C.3
Graphical interpretation of bipolar design rules.

(See color plate 7 in insert section)

TABLE 2C.4
Process parameters for a typical bipolar process[a]

Parameter	Typical	Tolerance[b]	Units
Ebers-Moll model parameters			
β_F (forward β)			
npn—vertical	100	50 to 200	
pnp—lateral			
(at I_C = 500 μA)	10	±20%	
(at I_C = 200 μA)	6	±20%	
β_R (reverse β)			
npn—vertical	1.5	±0.5	
pnp—lateral			
(at I_C = 500 μA)	5	±20%	
(at I_C = 200 μA)	3	±20%	
V_{AF} (forward Early voltage)			
npn—vertical	100	±30%	V
pnp—lateral	150	±30%	V
V_{AR} (reverse Early voltage)			
npn—vertical	150	±30%	V
pnp—lateral	150	±30%	V
J_S (saturation current density)			
npn—vertical	2.6×10^{-7}	−50%to + 100%	pA/μ^2
pnp—lateral	1.3×10^{-5}	−50%to + 100%	pA/μ emitter perimeter

TABLE 2C.4
(Continued)

Parameter	Typical	Tolerance[b]	Units
	Doping		
n^+ emitter	10^4	±30%	$10^{16}/cm^3$
p-base			
Surface	10^5	±20%	$10^{16}/cm^3$
Junction	1	±20%	$10^{16}/cm^3$
Epitaxial layer	0.3	±20%	$10^{16}/cm^3$
Substrate	0.08	±25%	$10^{16}/cm^3$
	Physical feature size		
Diffusion depth			
n+ emitter diffusion	1.3	±5%	μ
p-base diffusion	2.6	±5%	μ
p-resistive diffusion	0.3	±5%	μ
n-epitaxial layer	10.4	±5%	μ
n^+buried collector diffusion			
Into epitaxial	3.9	±5%	μ
Into substrate	7.8	±5%	μ
Oxide thickness			
Metal to epitaxial	1.4	±30%	μ
Metal to p-base	0.65	±30%	μ
Metal to n^+ emitter	0.4	±30%	μ
	Capacitances		
Metal to epitaxial	0.022	±30%	fF/μ^2
Metal to p-base diffusion	0.045	±30%	fF/μ^2
Metal to n^+ emitter diffusion	0.078	±30%	fF/μ^2
n^+ buried collector to substrate (junction, bottom)	0.062	±30%	fF/μ^2
Epitaxial to substrate (junction, bottom)	0.062	±30%	fF/μ^2
Epitaxial to substrate (junction, sidewall)	1.6	±30%	fF/μ perimeter
Epitaxial to p-base diffusion (junction, bottom)	0.14	±30%	fF/μ^2
Epitaxial to p-base diffusion (junction, sidewall)	7.9	±30%	fF/μ perimeter
p-base diffusion to n^+ emitter diffusion (junction, bottom)	0.78	±30%	fF/μ^2
p-base diffusion to n^+ emitter diffusion (junction, sidewall)	3.1	±30%	fF/μ perimeter

TABLE 2C.4
(Continued)

Parameter	Typical	Tolerance[b]	Units
Resistance and resistivity			
Substrate resistivity	16	±25%	Ω · cm
n^+ buried collector diffusion	17	±35%	Ω / □
Epitaxial layer	1.6	±20%	Ω · cm
p-base diffusion	160	±20%	Ω / □
p-resistive diffusion (optional)	1500	±40%	Ω / □
n^+ emitter diffusion	4.5	±30%	Ω / □
Metal	0.003		Ω / □
Contacts ($3\mu \times 3\mu$)	<4		Ω
Metal-n^+ emitter (contact plus series resistance to BE junction)	<1		Ω
Metal-p-base[c] (contact plus series resistance)	70		Ω
Metal-Epitaxial[d] (contact plus series resistance to BC junction)	120		Ω
Breakdown voltages, leakage currents, migration currents, and operating conditions			
Reverse breakdown voltages			
n^+ emitter to p-base	6.9	±50 mV	V
p-base to epitaxial	70	±10	V
Epitaxial to substrate	>80		V
Maximum operating voltage	40		V
Substrate leakage current	0.16		fA/μ^2
Maximum metal current density	0.8		mA/μ width
Maximum device operating temperature (design)	125		°C
Maximum device operating temperature (physical)	225		°C

[a] Process parameters based on the process of Tables 2C.1 and 2C.2.

[b] The tolerance is in terms of the absolute value of the parameter relating to processing variation from run to run. Matching characteristics on a die are much better. For example, β_F matching for identical devices on the same die to ±5% and J_S matching to ±5% are achievable.

[c] The base series resistance is strongly dependent on layout. Value given is for double base contact.

[d] The collector series resistance is strongly layout-dependent. It is primarily dependent on the vertical distance between contact and buried layer and buried layer to base.

TABLE 2C.5
SPICE model parameters of typical bipolar process

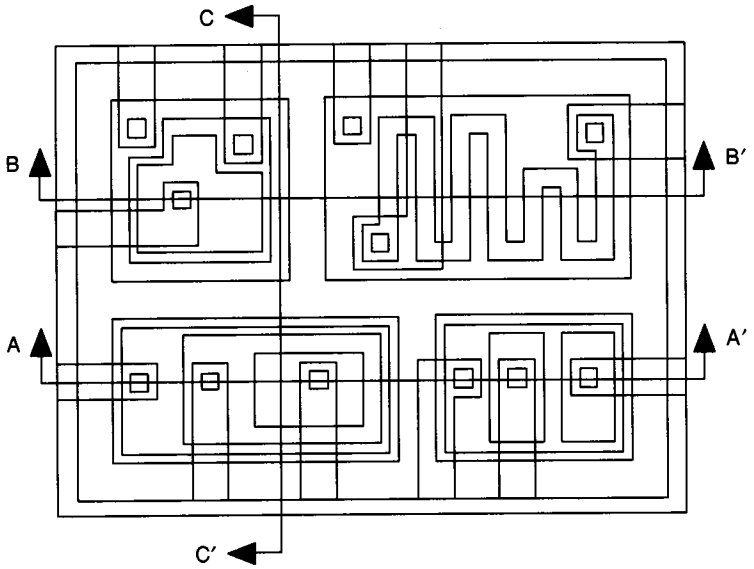
Parameter[a,b,c]	Vertical npn	Lateral pnp	Units
IS[c]	0.1	0.78	fA
BF	80	225	
NF	1	1	
VAF	100	150	V
IKF	100	0.1	mA
ISE	0.11	0.15	fA
NE	1.44	1.28	
BR	1.5		
NR	1	1	
VAR[b]	19	38	V
ISC		1.5	fA
NC	1.44	1.28	
RB	70	250	Ω
RE	1	4	Ω
RC	120	130	Ω
CJE	0.62	0.48	pF
VTE	0.69	0.65	V
MJE	0.33	0.40	
TF	0.45	40	ns
CJC	1.9	0.48	pF
VJC	0.65	0.65	V
MJC	0.4	0.4	
XCJC	0.5	0	
TR	22.5	2000	ns
CJS[d]	1.30	0	pF
VJS	0.49	0	pF
MJS	0.38	0	

[a] Parameters are defined in Chapters 3 and 4.

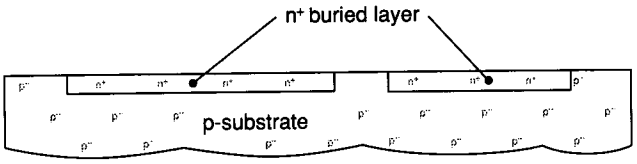
[b] Some of these Gummel-Poon parameters differ considerably from those given in Table 2C.4. They have been obtained from curve fitting and should give good results with computer simulations. The parameters of Table 2C.4 should be used for hand analysis.

[c] Parameters that are strongly area-dependent are based upon an npn emitter area of 390 μ^2 and perimeter of 80 μ, a base area of 2200 μ^2 and perimeter of 200 μ, and a collector area of 10,500 μ^2 and perimeter of 425 μ. The lateral pnp has rectangular collectors and emitters spaced 10 μ apart with areas of 230 μ^2 and perimeters of 60 μ. The base area of the pnp is 7400 μ^2 and the base perimeter is 345 μ.

[d] CJS is set to zero for the lateral transistor because it is essentially nonexistent. The parasitic capacitance from base to substrate, which totals 1.0 pF for this device, must be added externally to the BJT.

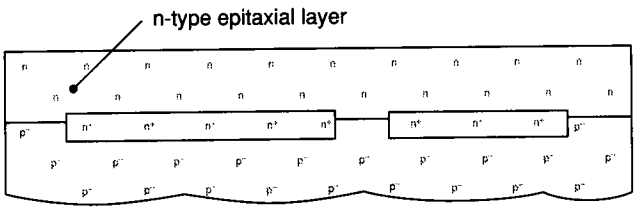


(*a*) Top View of Bipolar Die[a] (color version appears in Plate 4)



A-A′ Following Fabrication of Buried Layer

(*b*)

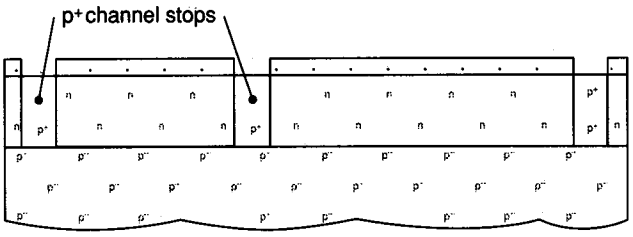


A-A′ Following Epitaxy

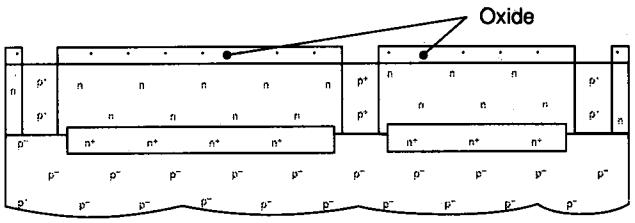
(*c*)

FIGURE 2C.1
Process description for the bipolar process.

[a] The base-emitter junction which must be reversed biased forms the capacitor in the upper left. A p-base diffused resistor is in the upper right. The vertical npn transistor and lateral pnp appear in the lower left and lower right respectively.

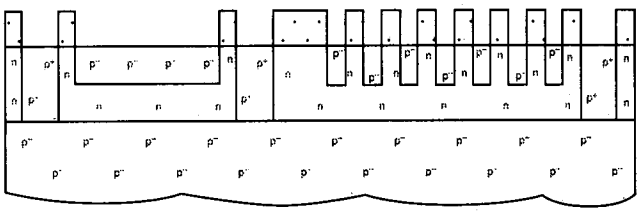


B-B′ Following Isolation Diffusion

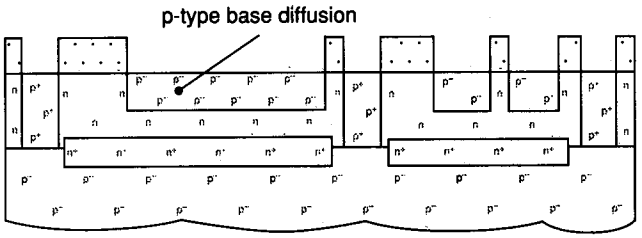


A-A′ Following Isolation Diffusion

(*d*)



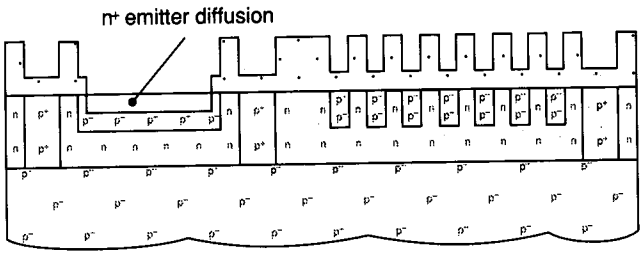
B-B′ Following p-type Base Diffusion



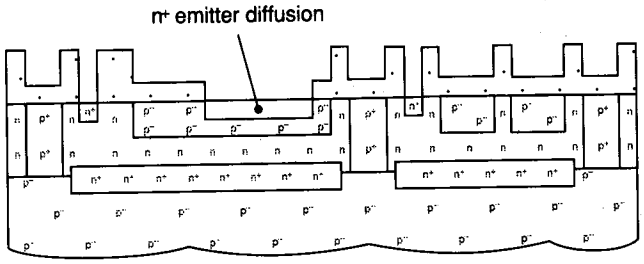
A-A′ Following p-type Base Diffusion

(*e*)

FIGURE 2C.1
(Continued)

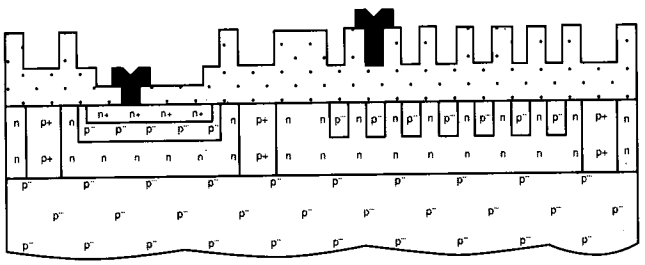


B-B′ Following n+ Emitter Diffusion

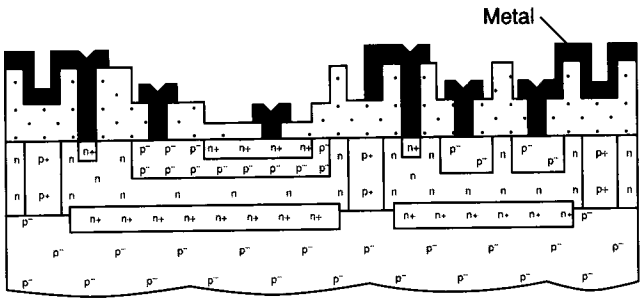


A-A′ Following n+Emitter Diffusion

(*f*)



B-B′ Following Metallization



A-A′ Following Metallization

(*g*)

FIGURE 2C.1
(Continued)

APPENDIX 2D
PROCESS CHARACTERIZATION OF A GENERIC THICK FILM PROCESS

TABLE 2D.1
Process scenario of major process steps in typical thick film process

1. Screen conductor material	(Screen 1)
2. Fire conductor material	
3. Screen resistive layer 1	(Screen 2)
4. Screen resistive layer 2	(Screen 3)
5. Screen resistive layer 3	(Screen 4)
6. Screen dielectric—first pass	(Screen 5)
7. Fire dielectric	
8. Screen dielectric—second pass	(Screen 6)
9. Fire dielectric	
10. Screen second conductor	(Screen 7)
11. Fire circuit	

TABLE 2D.2
Design rules for a typical thick film process

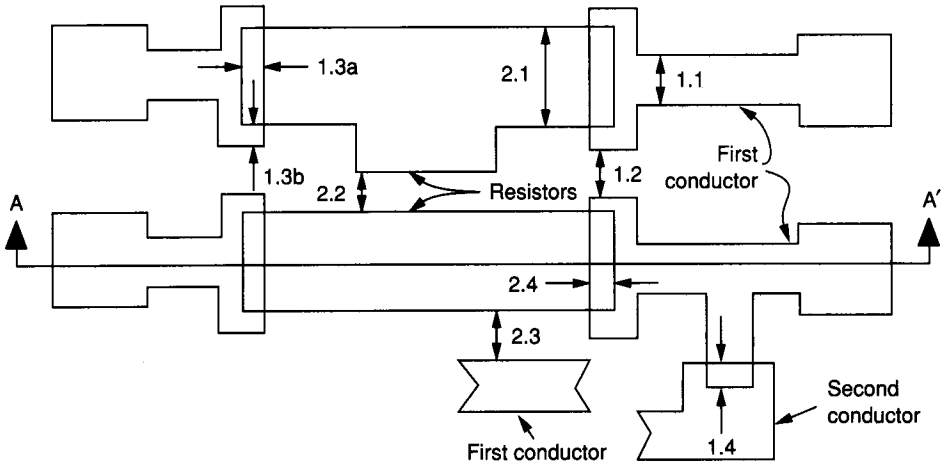
	Minimum dimension[a]
1. Conductor	
1.1 Width	2λ
1.2 Spacing	2λ
1.3 Metal larger than contacting resistor	
1.3*a* In direction of resistor	λ
1.3*b* Perpendicular to resistor	λ
1.4 Overlap of second contacting conductor	λ
1.5 Spacing to edge of substrate	2λ
2. Resistors	
2.1 Width or length	5λ
2.2 Spacing (same or different screenings)	2λ
2.3 Spacing to unconnected conductor	2λ
2.4 Overlap of conductor	λ
2.5 Spacing to resistor or conductor in trim region	
2.5*a* For laser trims	λ
2.5*b* For abrasive trims	4λ
2.6 Spacing to edge of substrate	2λ
3. Dielectric	
3.1 Dielectric larger than upper conductor (on top of lower conductor)	λ
4. Pad size	4λ

[a] λ is typically around 250 μ. Design rules depicted graphically in Fig. 2D.1.

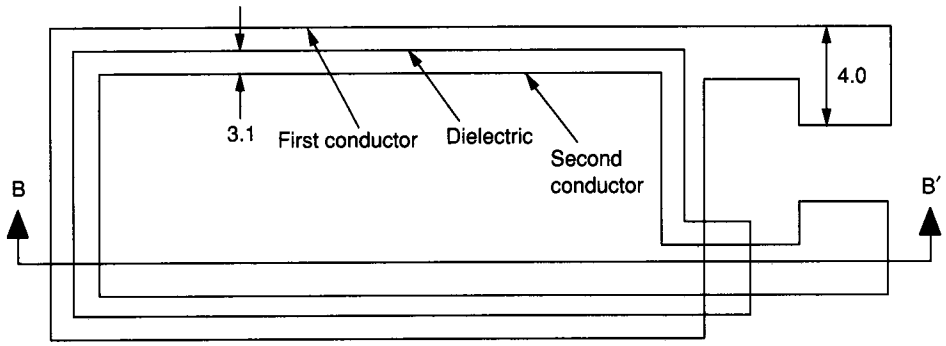
TABLE 2D.3
Process parameters for a typical thick film process

Parameter	Typical	Tolerance	Units
Conductor sheet resistance (after firing)	0.5	±40%	Ω/□
First resistor sheet resistance (after firing)	100	±30%	Ω/□
Second resistor sheet resistance (after firing)	1000	±30%	Ω/□
Third resistor sheet resistance (after firing)	10,000	±30%	Ω/□
Screen mesh size	200		per inch
Thickness after firing (resistors and conductors)	20	±20%	μ
Dielectric thickness	40	±30%	μ
Capacitance density[a]	0.8	±40%	pF/cm^2
TCR_1	−150	±20%	ppm/°C
TCR_2	+50	±20%	ppm/°C
TCR_3	200	±20%	ppm/°C
TCC	10	±10%	ppm/°C
VCR (all resistors)	≤ 60		ppm/V

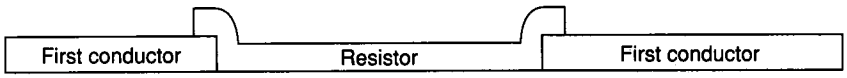
[a] The capacitance density is strongly a function of the characteristics of the material used for the dielectric.



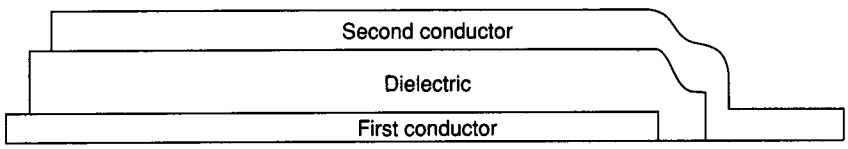
(*a*) Top view



(*b*) Top view



(*c*) Cross section A-A′



(*d*) Cross section B-B′

FIGURE 2D.1
Process description for the thick film process.

APPENDIX 2E
PROCESS CHARACTERIZATION OF A GENERIC THIN FILM PROCESS

TABLE 2E.1
Process scenario of major process steps in a typical resistive thin film process

1. Deposit resistive film	
2. Deposit conductive film	
3. Apply photoresist and pattern conductor	(Mask #1)
4. Etch conductor	
5. Apply photoresist and pattern resistor (and conductor)	(Mask #2)
6. Etch resistor	
7. Apply passivation	
8. Apply photoresist and pattern pad openings	(Mask #3)

TABLE 2E.2
Design rules for a typical thin film process

	Minimum dimensions
1. Conductor	
1.1 Width	200 μ
1.2 Spacing	200 μ
2. Resistor	
2.1 Width	20 μ
2.2 Spacing	20 μ
2.3 Spacing to conductor	200 μ
3. Pads	
3.1 Pad size	200 $\mu \times$ 200 μ
3.2 Passivation	180 $\mu \times$ 180 μ

Note: Design rules depicted graphically in Fig. 2E.1.

TABLE 2E.3
Process parameters for a typical thin film process

Parameter	Typical	Tolerance	Units
Conductor sheet resistance	0.25	±20%	Ω/□
Resistor sheet resistance	100	±20%	Ω/□
Temperature coefficient of resistance	100		ppm/°C
Voltage coefficient of resistance	5		ppm/V/mm length
Resistor matching (untrimmed)	±5%		
Conductor film thickness	1000		Å
Resistor film thickness	1000		Å

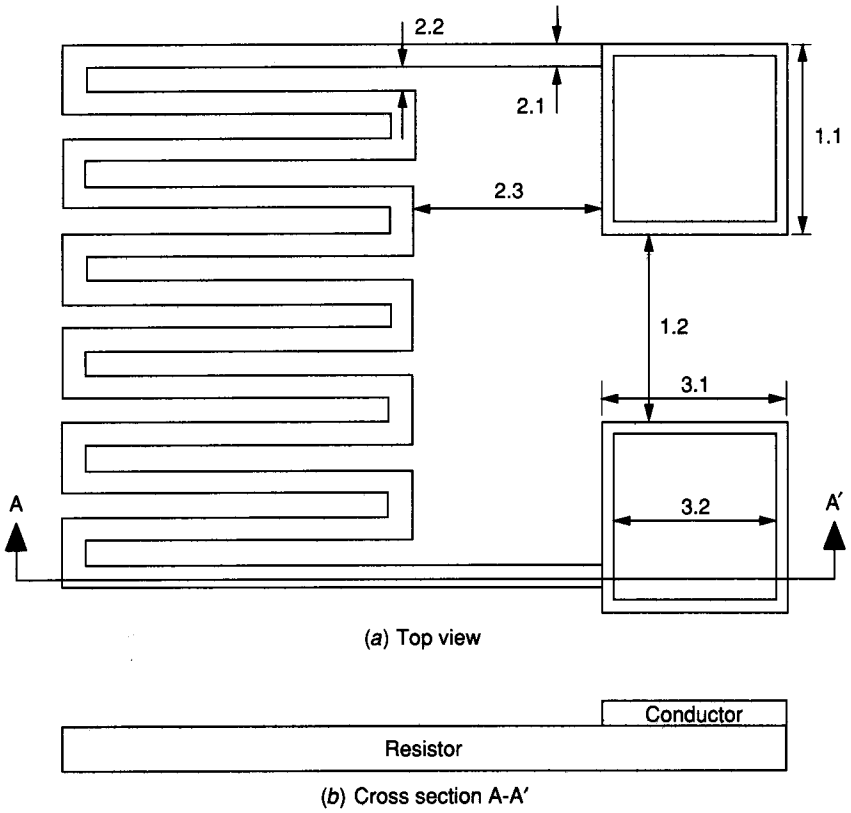


FIGURE 2E.1
Process description for the thin film process.